A new Relational Watermarking Scheme Resilient to Additive Attacks

Prof. R.Manjula
Associate Professor
School of Computing Sciences and Engineering
VIT University, Vellore
Tamil Nadu, India.

Nagarjuna Settipalli
M.Tech-CSE
School of Computing Sciences and Engineering
VIT University, Vellore
Tamil Nadu, India.

ABSTRACT
Strengthening the ownership rights on outsourced relational database is very important in today's internet environment. Especially where sensitive, valuable content is to be outsourced. Let us take an example of university database, weather data, stock market data, power consumption consumer behaviour data, and medical and scientific data. The increasing use of databases in applications beyond "behind–the- firewalls data processing" is creating a need for watermarking relational databases. Watermarking for relational data is made possible by fact that real data can very often tolerate a small amount of errors without any significant degradation with respect to their usability. In this paper, we present a mechanism that is resilient or insensitive to additive attacks, how to embed and detect watermark in relational database. In additive attack the attacker simply inserts his/her own watermark in original data. In our proposed system we can draw graphs and original ownership claim can be resolved by locating the overlapping regions of the two watermarks in which the bit values of the marks conflict and determining which owner's mark win. The attacker must have inserted the watermark later. Clearly having more marked tuples increases collisions and hence we can easily identify the owner of the data.

Index Terms
Watermarking, relational data, ownership rights, resilient, attacks, robust, secure.

1. INTRODUCTION
More than 700 years ago, watermarks are used in Italy to indicate the paper brand and the mill that produced it [13]. By the 18th century watermarks began to be used as anti-counterfeiting measures on money and other documents. The term watermark was introduced near the end of the 18th century. It was probably given because the marks resemble the effect of water on paper. The first example of technology similar to digital watermarking is a patent filled in 1954 by email hembrooke for identifying music works. In 1988, komastu and tominaga appear to be the first to use the term digital watermarking. About 1985, interest in digital watermarking began to mushroom [8]. The main aim of watermarking is to protect a certain data from unauthorized duplication and distribution by enabling provable ownership over the content [16].

More recently the focus of watermarking digital rights protection is shifting towards different data such as text, video, audio, software and relational data [19]. Watermarking embedding for relational data is made possible by the fact that real data can very often tolerate small amount of error without any significant degradation with respect to their usability. Detecting the watermark neither requires access the original data nor the watermark and the watermarking can be easily and efficiently maintained in the presence of insertion, updating and deletion. The increasing use of databases in application beyond "behind-firewall processing" is creating need for watermarking databases. Secure watermarking embedding requires that the embedded watermark must not be easily tampered with, forged or removed from the watermarked data. Imperceptible embedding means that the presence of watermark is unnoticeable in the data [3]. Basic characteristics of watermarking are robust, imperceptible, secure and reliable, low complexity, secure hiding place, payload, blind.

- Robustness: Robustness means the watermarking should be robust enough to handle any kind of situation [9], [2].
- Imperceptible: In some cases the watermarking is neither visible by human eyes nor hearable by human ears. It means it can be detected by special processing or special circuit only [1]. This means the watermark will not affect the original host data.
- Secure and Reliable: Watermark has unique correct signs marking every one, and thus to achieve the purpose of copyright protection.
- Low-Complexity: Low complexity algorithm will ensure effective and timely manner to watermarking embedding, detection and extraction.
- Secure Hiding Place: Watermark is embedded in correct place, and that will not be change the format of the original relational databases.
- Payload: The amount of information that can be stored in a watermarking.
- Blind: The watermark detection is blinded that means to extract the watermark from original data it requires neither the original data nor the watermark. So the watermark detection is blinded [14].

1.1 Applications of Watermarking
- Copyright protection: It is most prominent application. Embedded information about the owner to prevent others from claiming copyright [11].
- Copy protection: Embedded watermark to disallow unauthorized copying of original data.
- Content Authentication: Embedded a watermark to detect modifications to the host data.
- Traction Tracking: Embedded a watermark to convey information about the legal recipient of the data. This

is useful to monitor or trace back illegally produced copies of the data. This is usually referred as "fingerprinting".

- Broadcast Monitoring: Embedded a watermark in the original data and use automatic monitoring to verify whether data was broadcasted as agreed.

2. EXISTING SYSTEM

Database relations are updated or changed frequently, so there may be possibility of marks in the relational database can be removed or destroyed by malicious attacks. The attackers may try to change the original watermarked data or completely destroy the watermark. In this section we discus about possible malicious attacks [12], [15], [6], [20]. Till date only few types of attacks are overcomed they are bit attacks, randomization attacks, rounding attacks, subset attacks, insertion attacks, deletion attacks, geometric transforms, collusion attack.

2.1 Bit Attack

In bit attack [15] the attacker tries to destroy the watermark by simply updating some bits. If attacker can change all the bits then he can easily destroy the watermark. But one drawback of this type of attack to the attacker is he made data completely useless. Since each change can be considered an error. If more number of errors are there data completely useless. So this attack is overcomed.

2.2 Rounding Attack

In this the attacker tries to lose watermark contained in the numeric attribute by rounding to closest integer values. In this type of attack also the quality of data is degraded drastically, which means that the data is completely useless. Until he guesses the correct bit positions involved in the watermarking he may not be succeeded in his attack. Even if he guess the correct bits the more number of errors indicates that the data is useless.

2.3 Randomization Attack

Randomization attacker assigns random values to some number of bit positions. A zero out attack sets value of some number of bit positions to zero. If more number of bits are randomized the data becomes useless.

2.4 Subset Attack

Attacker may take a subset of tuples or attributes of a watermarked relation and hope that the watermark is destroyed or lost. If the attacker takes the many number of tuples or attributes the quality of data is degraded.

A sample snapshot of watermarked relational database table is shown in below Figure 1.

	A0	A1	A2	A3	A4	A5		A_{n-1}	A_n
Tuple1			■					■	
Tuple2		■		■			■		■
Tuple3	■				■	■			
Tuple4		■		■			■		■
Tuple5			■					■	
Tuple6			■					■	
Tuple7		■		■			■		■
Tuple8	■				■	■			
Tuple9		■		■			■		■
Tuple10			■					■	
……			■					■	
……		■		■			■		■
……	■				■	■			
$Tuple_{n-1}$		■		■			■		■
$Tuple_n$			■					■	

Figure 1. A snapshot of watermarked table

2.5 Mix and Match Attack

In mix and match attack the attacker may create his own relation by taking disjoint tuples from multiple relations containing similar information. In this case the attacker must not create the exact database as host database, and then we can easily identify who is the owner of the data and who is the attacker. Let us take an example attacker x takes f fraction of tuples from y's relation R and mixes them with tuples from other sources to create his relation S of the same size as R. Then the original data owner y able to his watermark in relation S by using the formula

f (n/r)+1/2(1-f)n/r>=T

2.6 Invertibility Attack

Attacker may launch an invertibility attack [5] to claim the ownership if he can successfully discover a fictitious watermark.

2.7 Collusion Attack

Why collusion attack [17] should be considered means if an attacker has access the more than one copy of data he/she can identify the or remove the watermark data by colluding them.

2.8 Insertion Attack

Attacker may decide to insert a tuple g to the data set D_W hopping to perturb the embedded watermark. The insertion of new tuples acts as additive noise to the embedded watermark. However the insertion of extra tuples may create many errors to the watermarked relational database. If more number of errors are there the database may not be useful. In addition to that addition of new tuples to the watermarked data may create a synchronization error.

2.9 Alteration Attack

In this attack, attacker alters the data value of g tuples. Here the attacker is faced with the challenge that altering the data may disturb the watermark. And the attacker does not have access rights to change the data set D, and thus he may easily violate the usability constraints and render the data useless.

2.10 Deletion Attack

In deletion attack the attacker intentionally deletes some g tuples from the marked data set. If the tuples are randomly deleted, then, on average, each partition loses g/m tuples. The watermark embedding which is discussed in [4] uses marker tuples to locate the start and end of data partitions. The embedded watermark is a stream of bits where the marker tuples identify the boundaries between the bits of the watermarked stream, which makes such marker-based watermarking technique [16] susceptible to watermark synchronization error. The successful deletion if tuples may create large number of errors in the data set. If more number of errors were there the database relation may be useless.

2.11 Mosaic Attack

In mosaic attack [7], [10] a data set is divided in to many small number of parts, this attack mostly possible in images and video and other multimedia data watermarking. After making the too small parts some small size data are deleted from the data set. This attack is also broken, because if attacker deletes the some amount of data from data set the data may not be useful, and the attacker may not claim the ownership of data. Because if a small amount of data can be deleted the quality of relation degraded drastically. In the same way the mosaic attack if the attacker deletes a small amount of data from the watermark is disturbed and the data also get damaged which is not useful for the attacker.

Table 1. Notations

f	Fraction of tuples
n	Number of tuples in the data set
T	Minimum number of correctly marked tuples needed for deletion
r	Fraction tuples marked
j	Watermark length
D_W	Watermark data set
W_D	Detected watermark set
m	Number of partitions in the relation
$P_0,\ldots,P_{m-1}$	Partition set
$a_{j-1},\ldots a_0$	Watermark bits
K_s	Secret key

All the above attacks are overcomed in earlier approaches of watermarked relational database system, multimedia watermarking system, blind pattern matching attack on watermarking system, digital watermarking technology. In all the above different types of attacks if the attacker tries to modify, insert, delete the some data, but in al the cases the data is useless. In our proposed system explain about new type of attack that is additive attack. The new type of attack that is additive attack is also discussed [18], [13] on xml data.

3. PROPOSED SYSTEM

Additive attack means the attacker inserts their own watermark over the original watermarked data of type multimedia data such as audio data, video data, image data, xml data, and relational data. And claim the "legal" ownership of data. This type of attack is different from insertion, deletion and updating attack where data is little bit changed and watermarked data may not be useful. But in additive attack the attacker does not change or delete the watermarked data he/she simply adds his/her own watermark. Since the watermarks inserted afterwards is able to overwrite the original former watermark in some overlapping regions. It results in the illegal copy more detected element than original one. In this attack the attacker simply takes the original data identifies at what percent owner embedded the watermark. By applying less than the host watermarked data he/she again apply the watermark. In our proposed system we can draw the sample graphs that show the overlapping regions, and differentiate the attacker and owner of the data. Let Q be the total number of marked attributes in the original database relation, R be the number attributes available for marking in the same original database relation, and let Z be the total number of watermarks added afterwards. The probability of having overlapping regions is defined as:

$$\begin{cases} 1-\prod_{i=0}^{R-1}\frac{(L-i)-M}{L-I}, & \text{if } M+R<L \\ 0, & \text{if } M+R>=L \end{cases}$$

We can calculate the mean of overlapping regions by using formula

Mean of the overlapping regions= L* Probability of collision in an element node =L*(Q/L)*(R/L).

Illegal parties may try to reduce the overlapping regions by using a low watermarking ratio such as 0.1% or 0.01%.

In our proposed system we can take different number of tuples in each time and we can draw the graphs. Figure 2 shows the probability and the mean of having overlapping regions when used 10,000 numbers of tuples in the database. We can take percentage of watermarks used in x-axis and probability of collision percentage in y-axis. By keep on increasing the watermarks ratio and number of tuples in the database relation

Case 1:

In the first case we can start the watermarking ratio 0.1, in this case we can identify the probability of collision. Here we can use 10,000 numbers of tuples in the database relation.

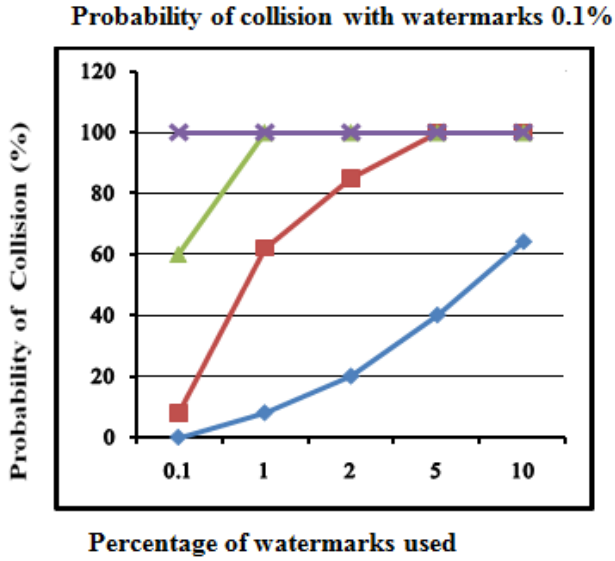

Figure 2. Probability of collision with 0.1% watermark ratio.

Case 2:

When we take 100,000 numbers of tuples in the database relation the graphs can be shown as:

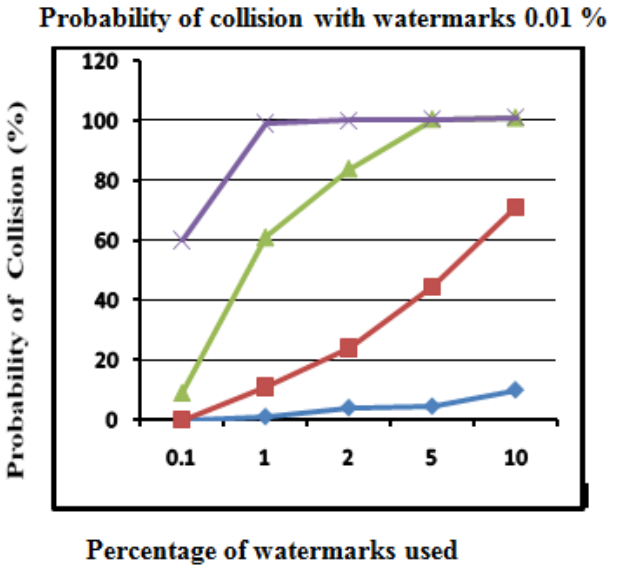

Figure 3. Probability of collision with 0.01% watermark ratio.

In the above case the percentage of watermarking is 0.01.

Case 3:

In tshis case we can take 1,000,000 number of tuples in the dtabase relation and we will draw the graph to find out overlapping regions with watermarking 0.1%.

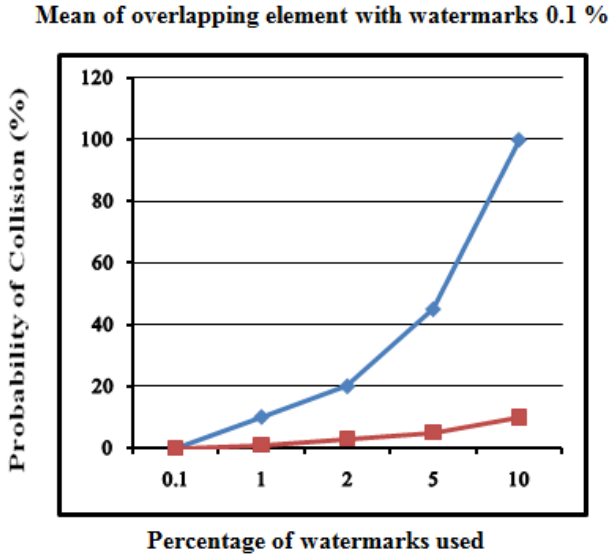

Figure 4. Mean of overlapping elements with 0.1% watermark ratio.

Case 4:

In last case we can use maximum number of tuples in the data base that is 10,000,000 and the percentage of watermarking is

0.01%. In this case we can identify the mean of overlapping regions. Here we can identify the mean of overlapping elements with watermarking ratio 0.01.

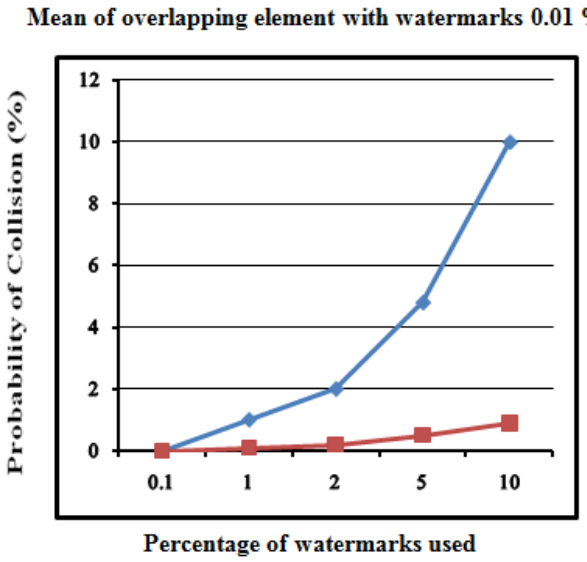

Figure 5. Mean of overlapping elements with 0.01% watermark ratio.

From the above graphs we can show that if the owner uses a 10% watermark ratio and the illegal party inserts watermarks with a 0.1% watermark ratio, the probability of the occurrence of overlapping regions is 65%. However, the mean of these overlapping regions is only 1 shown in figure 4. For a large relational document (N = 100,000), if the owner uses a lower watermark ratio of 2% and the illegal party inserts watermarks with a 0.1% watermark ratio, we can achieve a higher probability of overlapping (85%) and the mean of overlapping region 2. Figure 3 and 5 shows attackers using low watermarking ratio 0.01%.In this case, since the probability of overlapping region is low, in order to resist additive attacks with a very low watermark ratio, we can decrease the value of r such that when overlapping occurs, the collisions of watermarks are large enough to make an accurate decision.

4. APPLYING WATERMARKING TO RELATIONAL DATABASES

4.1 Watermark Embedding

Before embedding the watermark [4] in to original database first the original data is divided in to P_0, P_1,…,P_{m-1}. Where m-is the number of partitions in the database. A watermark is a set of j bits w=a_{j-1},…,a_0. The watermark length j is selected such that j<<m. The watermark bit a_i is embedded in partition P_k such that k mod j=i.

4.1.1 Watermark Embedding Procedure

Input: Data set D, Watermark bits w= {a_0,…,a_{j-1}}, Secret key k_s, number of partitions m.

Output: Watermarked data set D_w

Step 1: Partition the dataset D into P_0,…,P_{m-1}

Step 2: For each partition embedded a watermark bit such that i<-k mod j

End for loop

Step 3: Store watermarked data into D_w, return D_w.

4.2 Watermark Detection

The watermark detection is blinded that means it neither requires the knowledge of the original data nor is the watermark. The watermark detection starts from data partitions which is discussed in above section {P_0,P_1,…,P_{m-1}}. The watermark bit a_i extracted from partition P_k, such that k mod j=i, by using the watermarked dataset D_w secret key k_s and number of partitions m.

4.2.1 Watermark Detection Procedure

Input: Watermarked dataset D_w, number of partitions m, secret key k_s, watermark length j.
Output: Detected watermarked W_D
Set ones [0…j-1]<-0
Set zeros [0…j-1]<-0
Step 1: Get the partitions P_0,P_1…,P_{m-1}
Step 2: for i=0…m-1
I<-j mod l
If value greater than threshold
Ones[i]<-ones[i]+1;
else
zeros[i]<-zeros[i]+1;
end i;
end for;
Step 3: for i=0…m-1
If ones[j]greater than zeros[i]
Store 1 into W_D
else
store 0 into W_D
end if;
end for;
Step 4: return W_D

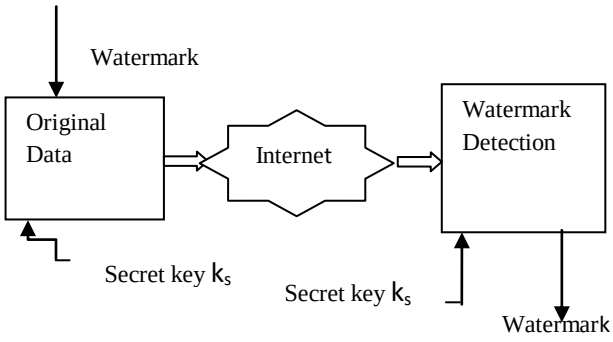

Figure 6. Watermark Embedding and Detecting

5. RELATED WORK

Ashraf Odeh and Ali Al-Haj classify the various types of attacks in watermarked relational database. In [4] paper they discussed about how to overcome the different types of attacks on relational database, and they proposed an effective database watermarking algorithm. RakeshAgrawal and Jerry Kiernan [15] also explained about various possible attacks in watermarking relational databases they are bit attacks, randomization attacks, rounding attacks, subset attack, mix and match attack, invertibility attack. RakeshAgrawal and jerry Kiernan proved that our watermarking technique is robust against all the above attacks they tested this algorithm on real world relational database. RaduSion [14] discussed how to boost up the ownership over categorical data he provided the limits to embedded the watermark in relational data. Ms. ArtiDeshpande and Mr. JayantGadge [3] proposed an algorithm for how to embedded the watermark in relational database and how to detect the watermark from original watermarked table, data partitioning algorithm.DarkoKirovski and Fabien A. P. Petitcolas [6] explained blind pattern matching attack on watermark, and different types of attacks on audio data, strength of watermark on relational database. In [16] RaduSion, Mikhail Atallah and Sunil Prabhakar discussed about rights protection for relational data, challenges of watermarking relational database, and optimization of watermark embedding, they introduces a one algorithm for embedding the watermark in relational database. And the primary key dependencies in relational data. Yanqun Zhang [19] explained what the basic characteristics of watermark and discussed theoretical model of watermarking. The closest to our technique is discussed in an effective approach for watermarking xml data [18], in this paper they discussed about what are the different types of attacks on watermarking relational database and explained about additive attack on xml data. How additive attack is applied by the attacker in various tags in xml data and how to identify those attacks is discussed in this paper. We are taken the same problem on relational database.

6. CONCLUSION

In this paper we discussed different types of attacks on relational database. We presented new watermarking technique, resilient to additive attack. In additive attack the attacker inserts his/her own watermark on original data and claim the ownership of data. We can draw the different graphs by taking various numbers of tuples in the table and watermarking ratio. Since the watermark inserted afterwards is able to overwrite the former watermarks in some overlapping regions, by calculating mean of overlapping regions and probability of collision with different watermarking ratio's we can differentiate the original owner of data and attacker. The probability of overlapping is very low in case of using low watermarking ratio. If the attacker applied watermark more the owner then data may not be useful. In existing system different types of attacks are discussed and overcomed by using different algorithms but not provide the exact solution to additive attacks. Our technique can easily identify the owner of the relational data.

7. FUTURE WORK

Our model does not prevent watermark once watermark applied but it can identify the correct owner of data. This means once watermark is applied to relational data the attacker again apply the watermark, this problem is not overcome by our technique. And if attacker applied more watermark than the host data the data can be useless and we can't rollback the damaged data.

8. REFERENCES

[1] Ajay Goel, "Improved Digital Watermarking Techniques and Data Embedding In Multimedia."Department of CSE Singhania University, Rajasthan, Rupesh Gupta Department of Mechanical Engineering, Singhania University Rajasthan, O.P.Sahu Department of ECE, N.I.T. Kurukshetra, Sheifali Gupta Department of ECE Singhania University, Rajasthan, India , 2010.

[2] Ali Al-Aaj and Ashrafbdeh, "Robust and Blind Watermarking of Relational Database Systems."Princess sumaya University for Technology AI-Jubeiha,Jordan, 2008.

[3] Ms. ArtiDeshpande, Mr. JayantGadge, "New Watermarking Technique for Relational Databases." Department of Computer Engineering, Thadomal Shahani Engineering College, Mumbai, ICETET-2009.

[4] Ashraf Odeh and Ali Al-Haj, "Watermarking Relational Database System." Arab Academy for Financial and Banking Sciences**,** 2Princess Sumaya University for Technology, Amman, JORDA, 2008.

[5] S. Craver, N.Memon, B.-L.Yeo, andM. M. Yeung. "Resolving rightful ownerships with invisible watermarking techniques: Limitations, attacks, and implications. "IEEE Journal of Selected Areas in Communications, 16(4):573–586, 1998.

[6] DarkoKirovski and Fabien A. P. Petitcolas, "Blind Pattern Matching Attack on Watermarking Systems. "IEEE Tractions on signal processing, Vol. 51, NO. 4, April 2003.

[7] Fabien A.P. Petitcolas, Ross J. Anderson, and Markus G. Kuhn, Fabien A.P. Petitcolas, Ross J. Anderson, and Markus G. Kuhn, "Attacks on Copyright Marking System."Notes in Computer Science, Portland, Oregon, USA, 14 17 April, 1998.

[8] Fernando Perez-Gonzalez and Juan R. Hernandez, "A Tutorial on Digital Watermarking." Dept. Tecnologıas de las Comunicaciones, ETSI Telecom., Universidad de Vigo, Spain, 1999.

[9] Frank Hartung and Bernd Girod, "Watermarking of Uncompressed and compressed video."Telecommunication Institute, University of Erlangen-Nurembery, 1998.

[10] Jonathan K. Su, Frank Hartung and Bernd Girod," Spread Spectrum Watermarking: Malicious Attacks and Counterattacks." Telecommunication Laboratory, University of Erlangen Nuremberg, Germany, 1999.

[11] S.P .Mahanty, "A Tutorial Review Report." Department of Electrical Engineering, Indian Institute of Sciences, Bangalore, India, 1999.

[12] Mohamed Shehab, ArifGhafoor, IEEE, Elisa Bertino, Fellow, "Watermarking Relational Databases Using Optimization-Based Techniques."IEEE, Vol.20, 2008.

[13] Podilchuk, C.I. and Delp., E.J, "Digital Watermarking: Algorithms and Applications. "IEEE Signal Processing Magazine, 2001.

[14] RaduSion, "Proving Ownership over Categorical Data. "Computer Sciences and the Center for Education and Research in Information Assurance and Security, Purdue University, USA, 2004.

[15] RakeshAgrawal, Jerry Kiernan, "Watermarking Relational Databases." IBM Almaden Research Center, china, 2002.

[16] R.Sion, M.Atallah, and S.Prabhkar, "Right Protection for Relational Data. "IEEE Trans. Knowledge and Data Engineering, Vol16 no.6, June 2004.

[17] VikasSaxena, J.P.Gupta, "Collision Attack Resilient Watermarking scheme for Colored Images Using DCT." IAENG, International journal of Computer Sciences, 2007.

[18] Wilfred Ng and Ho-Lam Lau, "Effective Approaches for Watermarking XML Data."Department of Computer

Science, the Hong Kong University of Science and Technology, Hong Kong, 2005**.**

[19] Yanqun Zhang, "Digital Water marking Technology: A Review."Department of Computer Science and Technology, China University of Mining and Technology,2009.

[20] Yingji u Li, Member, IEEE, VipinSwarup, and SushilJajodia, Senior Member, IEEE, "Fingerprinting Relational Databases: Schemes and Specialties." Vol no.2, March 2005.

Prof. R.Manjula received her B.E in Computer Science & Engineering from University of Vishwesvaraya and Engineering, Bangalore, Karnataka State, India in 1992 and M.E in Software Engineering from Anna University, Tamil Nadu, India in 2001. She is now working as Associate Professor and also as Ph.d Candidate affiliated with School of Computing Science and Engineering at Vellore Institute of Technology, Vellore, India. Her area of specialization includes Software Process modeling, Software Metrics, Software Metrics, Software Testing and Metrics, XML-Web Services and Service Oriented Architecture.

Nagarjuna.Settipalli received his B.Tech in Computer Science and Engineering from Newton's Institute of Engineering, Macherla, Andhra Pradesh State, India in 2008. He is now Pursuing his M.Tech at VIT University, Vellore, Tamil Nadu State, India. His interest includes watermarking, wireless network security, and he also developed a mobile application GSM Based Ticketing System for Airlines and a Cache Simulator as a miniproject.