Providing copyrights and Secured Transmission in a network through Watermarking

R. Lavanya[1,] Dr. M. Amanullah[2]

[1]: Assistant Professor, Department of Information Technology
Aalim Muhamed Salegh College of Engineering, Chennai-55
Email:r.lavanya@aalimec.ac.in

[2]: Associate Professor & HEAD, Department of Information Technology
Aalim Muhamed Salegh College of Engineering, Chennai-55
Email:m.amanullah@aalimec.ac.in

ABSTRACT

Proving possession rights on outsourced relational databases may be a crucial issue in today's internet-based application environments and in several content distribution applications. In this paper, a mechanism is provided for proof of possession supported the secure embedding of a sturdy unobservable watermark in relative information. We formulate the watermarking of relational databases as an affected optimisation downside and discuss economical techniques to resolve the optimisation downside and to handle the constraints. Our watermarking technique is resilient to watermark synchronization errors as a result of it uses a partitioning approach that does not need marker tuples.

Our approach overcomes a significant weakness in existing watermarking techniques. Watermark decoding is predicated on a threshold-based technique characterized by threshold based technique that minimizes the likelihood of decoding errors. In this paper a watermarking technique is implemented and shown by experimental results that the proposed technique is resilient to tuple deletion, alteration, and insertion attacks.

Keywords: Watermarking, Relational Databases, Encoding, Decoding

1. INTRODUCTION

The rapid growth of the Internet and related technologies has offered an unprecedented ability to access and redistribute digital contents. In such a context, enforcing data ownership is an important requirement, which requires articulated solutions, encompassing technical, organizational, and legal aspects.

Although we are still far from such comprehensive solutions, in the last years, watermarking techniques have emerged as a vital building block that plays a vital role in addressing the ownership problem. Such techniques permit the owner of the information to embed an imperceptible watermark into the data.

A watermark describes data which will be accustomed prove the possession of information like the owner, origin, or recipient of the content. Secure embedding requires that the embedded watermark must not be easily tampered with, forged, or removed from the watermarked data [3]. Imperceptible embedding implies that the presence of the watermark is unnoticeable within the knowledge.

Furthermore, the watermark detection is blinded, that is, it neither requires the knowledge of the original data nor the watermark. Watermarking techniques have been developed for video, images, audio, and text data [2,7,8], and also for software and natural language text [3,5]. To date, only a few approaches to the problem of watermarking relational data have been proposed [1,3]. These techniques, however, are not very resilient to watermark attacks.

In this paper, we present a watermarking technique for relational data that is highly resilient compared to these techniques. In explicit, our proposed technique is resilient to tuple deletion, alteration, and insertion attacks. Two techniques are formulated to solve formulated optimization problem based on genetic algorithms (GAs) and pattern search (PS) techniques. We present a data partitioning technique that does not depend on marker tuples to find the partitions and, thus, it is resilient to watermark synchronization errors. An Efficient technique for watermark detection that's based on optimum threshold. The optimum threshold is chosen by minimizing the likelihood of coding error.

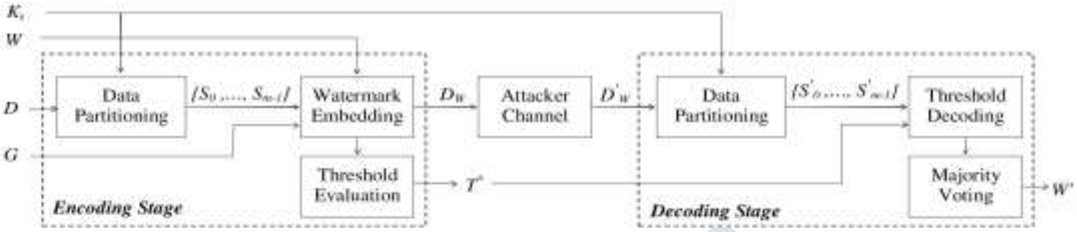


Fig. 1. Watermarking encoding and decoding process.

2. APPROACH OVERVIEW

A data set D is transformed into a watermarked version DW by applying a watermark encoding function that also takes as inputs secret key Ks only known to the copyright owner and a watermark W. Watermarking modifies the data. However, these modifications are controlled by providing usability constraints referred to by the set G. These constraints limit the amount alterations that can be performed on the data.

The watermark encoding is done by following steps: First the Data set is partitioned using the secret key and after partition watermarking embedding is done. A watermark bit is embedded in every partition by fixing the partition statistics whereas still confirming the usability constraints in G. This alteration is performed by solving a constrained optimization problem.
Finally, the bit embedding statistics are used to compute the optimal threshold T* that minimizes the probability of decoding error.

The watermarked version DW is delivered to the intended recipient. Then, it can suffer from unintentional distortions or attacks aimed at destroying the watermark information. Note that even intentional attacks are performed without any knowledge of Ks or D, since these are not publicly available.

The watermark decoding is split into 3 main steps: At first the Data set is partitioned using Data partitioning algorithm which was used already during encoding process then the data partitions are generated. The next step is Threshold-based decoding in which the statistics of every partition are evaluated, and the embedded bit is decoded using threshold-based scheme based on the optimal threshold T*. Finally, he watermark bits are decoded through a majority voting technique.

3. DATA PARTITIONING

In this section, we present the data partitioning algorithm that partitions the data set based on a secret key K. The data set D is a database relation with scheme D(P,A0, . . .,A_{v-1}), where P is the primary key attribute, A0, . . ., A_{v-1} are v attributes which are candidates for watermarking, and |D| is the number of tuples in D. The data set D is to be partitioned into m nonoverlapping partitions, namely, {S0,| . . ., | Sm-1}, such that each partition S_i contains on the average |D|/m tuples from the data set D. Partitions do not overlap, that is, for any two partitions S_i and S_j such that i ≠ j, we have Si n Sj = {}. For each tuple r ϵ D, the data partitioning algorithm computes a MAC, which is considered to be secure and is given by H(Ks||H(r.P||Ks)), where r.P is the primary key of the tuple r, H() is a secure hash function, and k is the concatenation operator. Using the computed MAC tuples are assigned to partitions. For a tuple r, its partition assignment is given by **partition(r)= H(Ks||H(r.P||Ks)) mod m.**

Using the property that secure hash functions generate uniformly distributed message digests this partitioning technique.. Furthermore, an attacker cannot predict the tuples-to partition assignment without the knowledge of the secret key Ks and the number of partitions m, which are kept secret. our technique can be easily extended to handle cases when the relation has no primary key.

Data Partitioning algorithm is shown below:

```
        S0,…,sm-1  ⟵  { }
For each tuple r Є D
        Partitions( r ) =H (Ks || H (r.P||Ks))
                                        mod m
        Insert r into S partition( r )
        return S0,…,sm-1.
```

Our data partitioning algorithm does not rely on special marker tuples for the selection of data partitions, which makes it resilient to watermark synchronization attacks caused by tuple deletion and tuple insertion.

4. WATERMARK EMBEDDING

In this section, we describe the watermark embedding algorithm by formalizing the bit encoding as a constrained optimization problem. Then, we propose a GA and a PS technique that can be used to efficiently solve such optimization problem.

Our watermarking technique is able to handle tuples with multiple attributes, as we will discuss in Section 6. However, to simplify the following discussion, we assume the tuples in a partition Si contain a single numeric attribute. In such a case each partition, Si can be represented as a numeric data vector Si =[si1, . . . , sin] ϵ Rn.

4.1 Single Bit Encoding

The bit encoding algorithm rule embeds bit bi within the partition Si if |Si| is bigger than ξ.
The value ξ. of represents the minimum partition size.

The maximize and minimize within the bit coding algorithm rule optimize the concealment perform θγ(Si +Δi*) subject to the constraints in Gi. The maximization and step-down answer statistics square measure recorded for every coding step in Xmax, Xmin, respectively. These statistics square measure accustomed work out best decipherment parameters.

4.2. Genetic Algorithm

A GA is a search technique that is based on the principles of natural selection or survival of the fittest, the GA uses the objective function directly in the search. The GA searches the solution space by maintaining a population of potential solutions.

Then, by victimization evolving operations like crossover mutation, and selection, the GA creates successive generations of solutions that evolve and inherit the positive characteristics of their parents and
thus gradually approach optimal or near-optimal solutions. By using the objective function directly in the search, GAs can be effectively applied in non convex, highly nonlinear, complex problems. GAs has been frequently used to solve combinatorial optimization problems and nonlinear problems with complicated constraints or non differentiable objective functions.

4.3. Pattern Search technique

PS methods are a class of direct search methods for nonlinear optimization. PS methods [4], have been widely used because of their implicitly and the fact that they work well in practice on a variety of problems. More recently, they are provably convergent . PS starts at an initial point and samples the objective function at a predetermined pattern of points centered about that point with the goal of producing a new better iterate.

4.4 Watermark Embedding Algorithm

A watermark is a set of l bits W = b_{l-1}, . . . , b_0 that are to be embedded in the data partitions {S0,| . . . ,|Sm_1}. To enable multiple embeddings of the watermark in the data set, the watermark length l is selected such that l<<m. The watermark embedding algorithm embeds a bit bi in partition Sk such that k mod l = i. This technique ensures that each watermark bit is embedded m/l times in the data set D. The watermark embedding algorithm generates the partitions by calling get partitions, then for each partition S_k, a watermark bit bi is encoded by using the single bit encoding algorithm (encode single bit) that was discussed in the previous sections. The generated altered partition S^{W}_{k} is inserted into watermarked data set D_W. Statistics (X_{max},X_{min}) are collected after each bit embedding and are used by the get optimal threshold algorithm to compute the optimal decoding threshold.

5. DECODING THRESHOLD EVALUATION

In the previous sections, we discussed the bit encoding technique that embeds a watermark bit b_i in a partition S_i to generate a watermarked partition $S^{W}{}_{i}$. In this section, we discuss the bit decoding technique that is used to extract the embedded watermark bit b_i from the partition S^{W}_{i}. The bit decoding technique is based on an optimal Threshold T* that minimizes the probability of decoding error. Presented with the data partition $S^{W}{}_{i}$, the bit decoding technique computes the hiding function $\theta\gamma(S^{W}_{i})$ and compares it to the optimal decoding threshold T* to decode the embedded bit bi. If $\theta\gamma(S^{W}_{i})$ is greater than T*, then the decoded bit is 1; otherwise, the decoded bit is 0. The decoding technique computes the normalized tail count of S^{W}_{i} by computing the reference ref and by counting the number of entries in S^{W}_{i} that are greater than ref. Then, the computed normalized tail count is compared to T*, see Fig. 2.

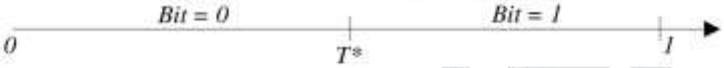
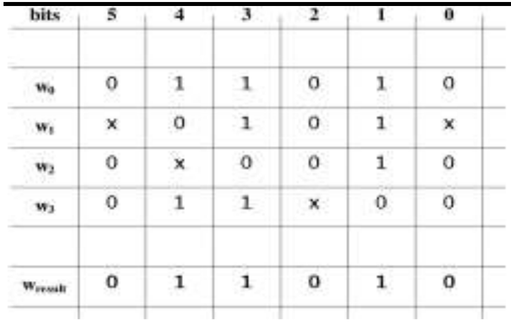


Fig. 2. Threshold-based decoding scheme.

The selection of the optimal T* is based on the collected output statistics of the watermark embedding algorithm. The optimal threshold T* minimizes the probability of decoding error and thus enhances the strength of the embedded watermark by increasing the chances of successful decoding.

6. WATERMARK DETECTION

In this section, we discuss the watermark detection algorithm that extracts the embedded watermark using the secret parameters including K_s,m, ξ, c, and T. The algorithm starts by generating the data partitions {S0,|. . ,| Sm_1} using the watermarked data set D_W, the secret key Ks, and the number of partitions m as input to the data partitioning algorithm discussed in Section 3. Each partition encodes a single watermark bit; to extract the embedded bit, we use the threshold decoding scheme based on the optimal threshold T* that minimizes the probability of decoding error, as discussed in Section 5. If the partition size is smaller than **ξ,** the bit is decoded as an erasure; otherwise, it is decoded using the threshold scheme.



bits	5	4	3	2	1	0
w_0	0	1	1	0	1	0
w_1	x	0	1	0	1	x
w_2	0	x	0	0	1	0
w_3	0	1	1	x	0	0
W_{result}	0	1	1	0	1	0

Fig. 3. An example illustrates the majority bit matching decoding algorithm for a watermark W = 011010, with "_" representing the erasures.

As the watermark W= bi-1, . . . , b0 is embedded several times in the data set, each watermark bit is extracted several times, where for a bit bi, it is extracted from partition Sk, where k mod l = i. The extracted bits are decoded using the majority voting technique, which is used in the decoding of repetition error correcting codes. Each bit bi is extracted m/l times so it represents a (m/l)-fold repetition code.

In case of a relation with multiple attributes, the watermark resilience can be increased by embedding the watermark in multiple attributes. The use of multiple attributes enables the multiple embedding of watermark bits many times in each partition, such embedding can be considered as an inner many-fold repetition code. For decoding purposes, the statistics Xmax and Xmin are collected for each attribute separately. The optimal threshold is computed for each attribute using the collected statistics to minimize the probability of decoding error. In the decoding phase, the watermark is extracted from each of the each attributes using the watermark detection algorithm, and then majority voting is used.

7. ATTACKER MODEL

In this section, we discuss the attacker model and the possible malicious attacks that can be performed. Assume that Alice is the owner of the data set D and has marked D by using a watermark W to generate a watermarked data set DW. The aggressor Mallory will perform many forms of attacks within the hope of corrupting or maybe deleting the embedded watermark. A robust watermarking technique should be ready to survive all such attacks. We assume that Mallory has no access to the initial information set D and doesn't grasp any of the key data employed in the embedding of the watermark, including the secret key Ks, the secret number of partitions m, the secret constant c, the optimization parameters, and the optimal decoding threshold T*. Given these assumptions, Mallory cannot generate the data partitions . . . , because this requires the knowledge of both the secret key Ks and the number of partitions m. We classify the attacks preformed by Mallory into 3 varieties, namely, deletion, alteration, and insertion attacks.

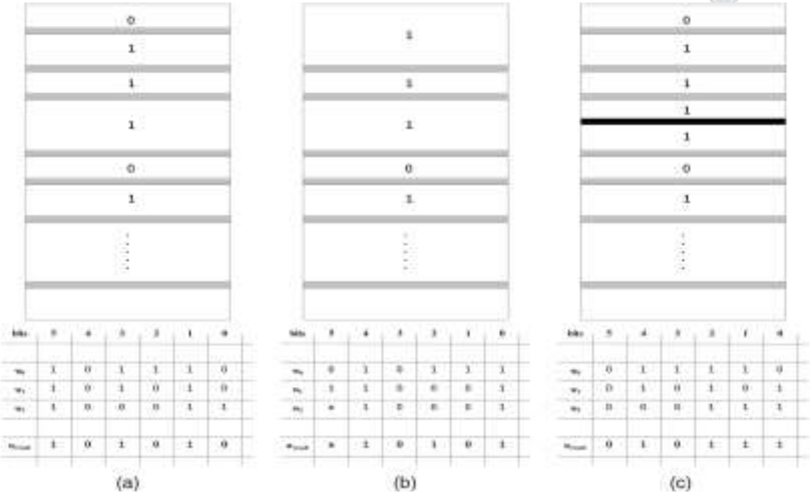
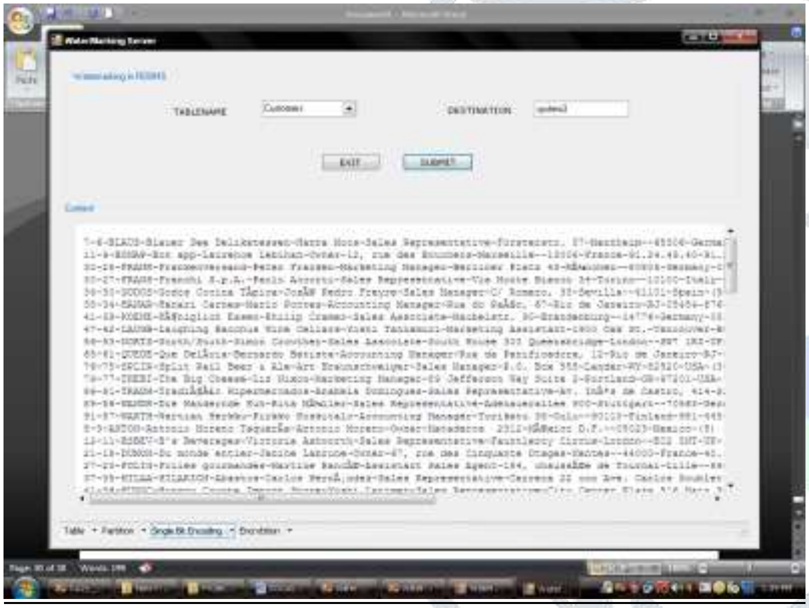


Fig. 4 Watermarked data set subject to the deletion and insertion attacks, respectively, and their corresponding majority voting maps. Gray-shaded cells represent the original marker tuples, and the black cells represent the added marker tuples. (a) Watermarked data set. (b) After deletion attack(c) After insertion attack.

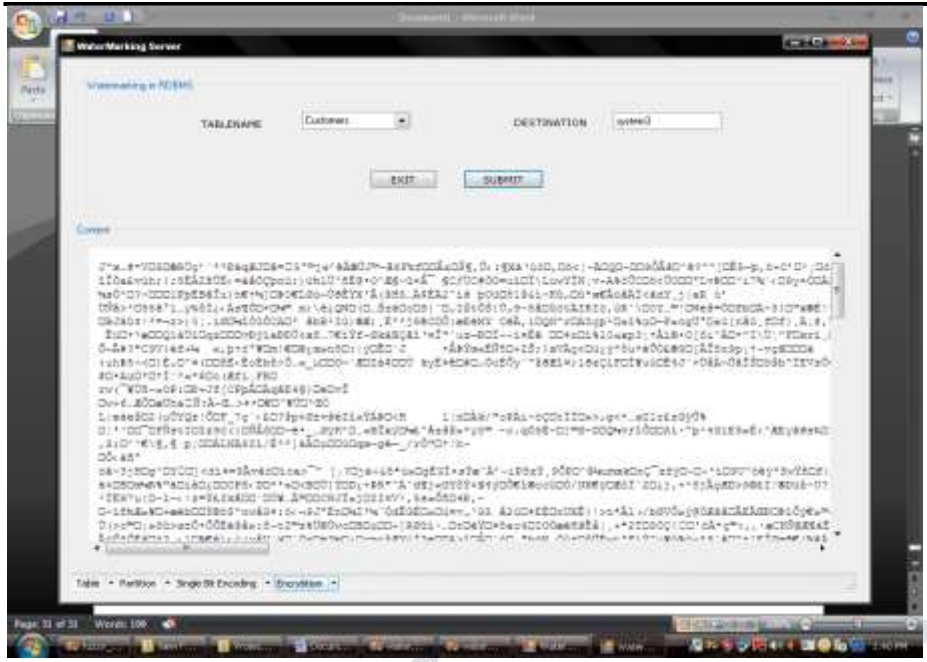
8. RESULTS AND DISCUSSIONS

In this proposed system, we have presented a resilient watermarking technique for relational data that embeds watermark bits in the data statistics. The Problem of watermarking was formulated as a constrained optimisation problem that maximizes or minimizes a hiding function based on the bit to be embedded. GA and PS techniques were utilized to unravel the planned optimisation problem and to handle the constraints. Furthermore, we presented a data partitioning technique that does not depend on special marker tuples to locate the partitions and proved its resilience to watermark synchronization errors. We developed an economical threshold-based technique for watermark detection that is based on an optimal threshold that minimizes likelihood of decipherment error. In decoding phase multiple attributes used and repeated watermark embedding done and majority voting technique used to improve watermark resilience.



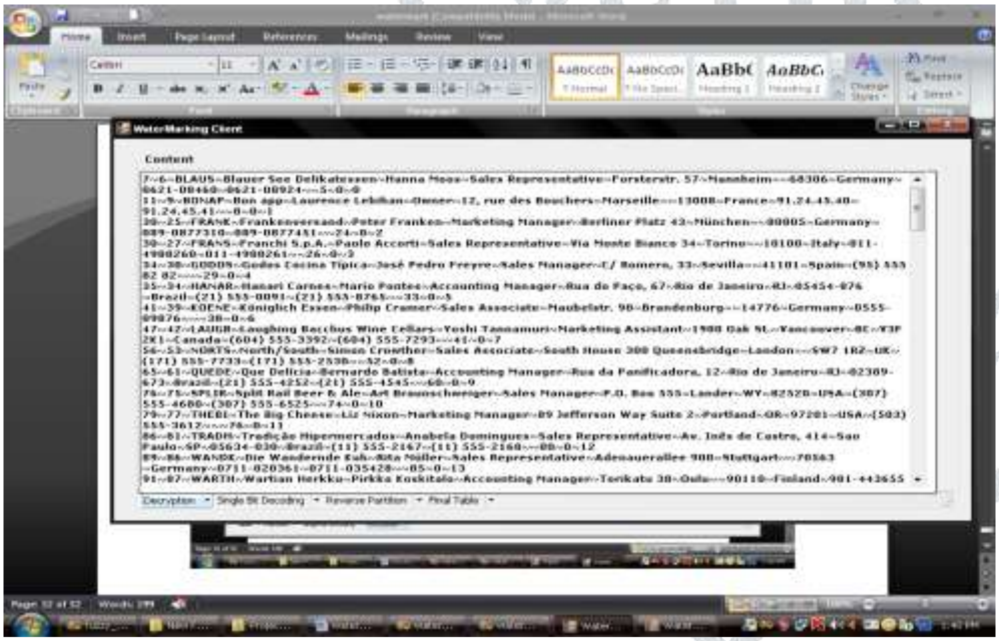
```
7-6-BLAUS-Blauer See Delikatessen-Hanna Moos-Sales Representative-Forsterstr. 57-Mannheim--68306-Germa
11-8-BONAP-Bon app-Laurence Lebihan-Owner-12, rue des Bouchers-Marseille--13008-France-91.24.45.40-91.
32-28-FRANK-Frankenversand-Peter Franken-Marketing Manager-Berliner Platz 43-MÃ¼nchen--80805-Germany-0
33-27-FRANS-Franchi S.p.A.-Paolo Accorti-Sales Representative-Via Monte Bianco 34-Torino--10100-Italy-
36-30-GODOS-Godos Cocina TÃpica-JosÃ© Pedro Freyre-Sales Manager-C/ Romero, 33-Sevilla--41101-Spain-(9
38-34-HANAR-Hanari Carnes-Mario Pontes-Accounting Manager-Rua do PaÃ§o, 67-Rio de Janeiro-RJ-05454-876-
41-39-KOENE-KÃ¶niglich Essen-Philip Cramer-Sales Associate-Maubelstr. 90-Brandenburg--14776-Germany-05
47-42-LAUGB-Laughing Bacchus Wine Cellars-Yoshi Tannamuri-Marketing Assistant-1900 Oak St.-Vancouver-B
56-53-NORTS-North/South-Simon Crowther-Sales Associate-South House 300 Queensbridge-London--SW7 1RZ-UK
65-61-QUEDE-Que DelÃcia-Bernardo Batista-Accounting Manager-Rua da Panificadora, 12-Rio de Janeiro-RJ-
78-75-SPLIR-Split Rail Beer & Ale-Art Braunschweiger-Sales Manager-P.O. Box 555-Lander-WY-82520-USA-(3
79-77-THEBI-The Big Cheese-Liz Nixon-Marketing Manager-89 Jefferson Way Suite 2-Portland-OR-97201-USA-
86-81-TRADH-TradiÃ§Ã£o Hipermercados-Anabela Domingues-Sales Representative-Av. InÃªs de Castro, 414-S
89-86-WANDK-Die Wandernde Kuh-Rita MÃ¼ller-Sales Representative-Adenauerallee 900-Stuttgart--70563-Ger
91-87-WARTH-Wartian Herkku-Pirkko Koskitalo-Accounting Manager-Torikatu 38-Oulu--90110-Finland-981-443
5-3-ANTON-Antonio Moreno TaquerÃa-Antonio Moreno-Owner-Mataderos  2312-MÃ©xico D.F.--05023-Mexico-(5)
13-11-BSBEV-B's Beverages-Victoria Ashworth-Sales Representative-Fauntleroy Circus-London--EC2 5NT-UK-
21-18-DUMON-Du monde entier-Janine Labrune-Owner-67, rue des Cinquante Otages-Nantes--44000-France-40.
27-25-FOLIG-Folies gourmandes-Martine RancÃ©-Assistant Sales Agent-184, chaussÃ©e de Tournai-Lille--59
37-35-HILAA-HILARION-Abastos-Carlos HernÃ¡ndez-Sales Representative-Carrera 22 con Ave. Carlos Soublet
41-38-HUNGC-Hungry Coyote Import Store-Yoshi Latimer-Sales Representative-City Center Plaza 516 Main S
```

Fig.5a). Bit Encoding in Relational Databases – embeds watermark bits in data statistics of Relational Data.



[illegible]

Fig.5b).Encrypt contents of table in Relational Databases before transmission of data.



```
7~6~BLAUS~Blauer See Delikatessen~Hanna Moos~Sales Representative~Forsterstr. 57~Mannheim~~68306~Germany~
0621-08460~0621-08924~~~5~0~0
11~9~BONAP~Bon app~Laurence Lebihan~Owner~12, rue des Bouchers~Marseille~~13008~France~91.24.45.40~
91.24.45.41~~~8~0~1
30~25~FRANK~Frankenversand~Peter Franken~Marketing Manager~Berliner Platz 43~München~~80805~Germany~
089-0877310~089-0877451~~~24~0~2
30~27~FRANS~Franchi S.p.A.~Paolo Accorti~Sales Representative~Via Monte Bianco 34~Torino~~10100~Italy~011-
4988260~011-4988261~~~26~0~3
34~30~GODOS~Godos Cocina Típica~José Pedro Freyre~Sales Manager~C/ Romero, 33~Sevilla~~41101~Spain~(95) 555
82 82~~~~29~0~4
35~34~HANAR~Hanari Carnes~Mario Pontes~Accounting Manager~Rua do Paço, 67~Rio de Janeiro~RJ~05454-876
~Brazil~(21) 555-0091~(21) 555-8765~~~33~0~5
41~39~KOENE~Königlich Essen~Philip Cramer~Sales Associate~Maubelstr. 90~Brandenburg~~14776~Germany~0555-
09876~~~~38~0~6
47~42~LAUGB~Laughing Bacchus Wine Cellars~Yoshi Tannamuri~Marketing Assistant~1900 Oak St.~Vancouver~BC~V3F
2K1~Canada~(604) 555-3392~(604) 555-7293~~~41~0~7
56~53~NORTS~North/South~Simon Crowther~Sales Associate~South House 300 Queensbridge~London~~SW7 1RZ~UK~
(171) 555-7733~(171) 555-2530~~~52~0~8
65~61~QUEDE~Que Delícia~Bernardo Batista~Accounting Manager~Rua da Panificadora, 12~Rio de Janeiro~RJ~02389-
673~Brazil~(21) 555-4252~(21) 555-4545~~~60~0~9
76~75~SPLIR~Split Rail Beer & Ale~Art Braunschweiger~Sales Manager~P.O. Box 555~Lander~WY~82520~USA~(307)
555-4680~(307) 555-6525~~~74~0~10
79~77~THEBI~The Big Cheese~Liz Nixon~Marketing Manager~89 Jefferson Way Suite 2~Portland~OR~97201~USA~(503)
555-3612~~~~76~0~11
86~81~TRADH~Tradição Hipermercados~Anabela Domingues~Sales Representative~Av. Inês de Castro, 414~Sao
Paulo~SP~05634-030~Brazil~(11) 555-2167~(11) 555-2168~~~80~0~12
89~86~WANDK~Die Wandernde Kuh~Rita Müller~Sales Representative~Adenauerallee 900~Stuttgart~~70563
~Germany~0711-020361~0711-035428~~~85~0~13
91~87~WARTH~Wartian Herkku~Pirkko Koskitalo~Accounting Manager~Torikatu 38~Oulu~~90110~Finland~981-443655
```

Fig.5c).Decrypt contents of table in Relational Databases after receiving of data.

9. CONCLUSION

In this paper, the watermark resilience was improved by the repeated embedding of the watermark and using majority voting technique in the watermark decoding phase. Moreover, the watermark resilience was improved by using multiple attributes. A proof of implementation of our watermarking technique was utilized to conduct experiments using artificial and real-world knowledge of data's. A comparison our watermarking technique with previously posed techniques shows the superiority of our technique to deletion, alteration, and insertion attacks.

In future work, we plan to apply some error correction code on decoding stages. We also planned to apply some security driven approach in relational database watermarking. We plan to propose a technique to detect and localize alterations made to a database relation with categorical attributes.

REFERENCES

[1] R. Agrawal and J. Kiernan, “Watermarking Relational Databases,” Proc. 28th Int’l Conf. Very Large Data Bases, 2002.

[2] M. Atallah and S. Lonardi, “Authentication of LZ-77 Compressed Data,” Proc. ACM Symp. Applied Computing, 2003.

[3] R. Wolfgang, C. Podilchuk, and E. Delp, “Perceptual Watermarks for Digital Images and Video,” Proc. IEEE, vol. 87, pp. 1108-1126,July 1999.

[4] G. Box, “Evolutionary Operation: A Method for Increasing Industrial Productivity,” Applied Statistics, vol. 6, no. 2, pp. 81-101, 1957.

[5] R. Sion, M. Atallah, and S. Prabhakar, “Rights Protection for Relational Data,” IEEE Trans. Knowledge and Data Eng., vol. 16, no. 6, June 2004.

[6] D. Gross-Amblard, “Query-Preserving Watermarking of Relational Databases and XML Documents,” Proc. 22nd ACM. Principles of Database Systems (PODS ’03), pp. 191-201, 2003.

[7] F. Hartung and M. Kutter, “Multimedia Watermarking Techniques,” Proc. IEEE, vol. 87, no. 7, pp. 1079-1107, July 1999.

[8] G. Langelaar, I. Setyawan, and R. Lagendijk, “Watermarking Digital Image and Video Data: A State-of-the-Art Overview,” vol. 17, no. 5, pp. 20-46, Sept. 2000.

[9] F. Petitcolas, R. Anderson, and M. Kuhn, “Attacks on Copyright Marking Systems,” LNCS, vol. 1525, pp. 218-238, Apr. 1998.