HAL open science

Preventing Additive Attacks to Relational Database Watermarking

Maikel Gort, Martina Olliaro, Claudia Feregrino-Uribe, Agostino Cortesi

To cite this version:

Maikel Gort, Martina Olliaro, Claudia Feregrino-Uribe, Agostino Cortesi. Preventing Additive Attacks to Relational Database Watermarking. 13th International Conference on Research and Practical Issues of Enterprise Information Systems (CONFENIS), Dec 2019, Prague, Czech Republic. pp.131-140, 10.1007/978-3-030-37632-1_12 . hal-03408432

HAL Id: hal-03408432

https://inria.hal.science/hal-03408432v1

Submitted on 29 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Preventing Additive Attacks to Relational Database Watermarking

Maikel Lázaro Pérez Gort[1], Martina Olliaro[2,3],
Claudia Feregrino-Uribe[1], and Agostino Cortesi[2]

[1] National Institute of Astrophysics, Optics and Electronics, Puebla, Mexico
{mlazaro2002es,cferegrino}@inaoep.mx
[2] Ca' Foscari University, Venice, Italy
{martina.olliaro,cortesi}@unive.it
[3] Masaryk University, Brno, Czech Republic

Abstract. False ownership claims are carried on through additive and invertibility attacks and, as far as we know, current relational watermarking techniques are not always able to solve the ownership doubts raising from the latter attacks. In this paper, we focus on additive attacks. We extend a conventional image-based relational data watermarking scheme by creating a non-colluded backup of the data owner marks, the so-called secondary marks positions. The technique we propose is able to identify the data owner beyond any doubt.

Keywords: Additive attack, False ownership claim, Relational data, Robust watermarking.

1 Introduction

Internet has made publicly available digital data on large scale, allowing users to fraudulently claim data ownership. In the 90s, digital watermarking techniques were developed to protect ownership rights of multimedia assets (i.e., images, audio, video, and texts), where a mark is permanently and unalterably placed into the latter. To overcome watermarking and counterfeit data intellectual property, several attacks have been conceived, and efforts in developing effective digital copyright protection mechanisms have been carried out in response. Invisible watermarking techniques increase the likelihood of successful prosecution once a theft has occurred [4]. Robust watermarking schemes are able to survive against watermark (WM) removal attempts and data manipulations (both malicious and benign). Finally, non-invertible watermarking techniques tackle those attacks, which makes possible multiple data ownership claims [6].

At the beginning of the 2000s, watermarking techniques were extended to relational data. As well as multimedia data watermarking, relational data watermarking techniques too had to deal with several attacks attempting both to remove the WM and to carry out false ownership claims [9]. Attacks attempting to raise doubts about data

ownership are called additive and invertibility attacks. According to [9], an additive attack is carried out when a malicious user adds his own WM to a watermarked relation and try to claim his ownership. On the other hand, an invertibility attack occurs when a malicious user is able to find a fictitious WM which is in fact a random occurrence from a watermarked relation.

This paper is focused on additive attacks. On it, we first discuss the basics and limitations of previous relational data watermarking techniques dealing with false claims of ownership carried out through additive attacks. Then we extend the image-based relational watermarking scheme presented in [7] by creating a non-colluded backup of the data owner's marks, the so-called secondary marks positions. The latter allows us to restore the owner's WM to determine the rightful data owner in case of been applied additive attacks over the protected data. Finally, we provide experimental results validating the proposed technique.

The rest of this paper is organized as follows. Section 2 discusses preliminaries about watermarking techniques for relational data, particularly the schemes created to deal with additive attacks. Section 3 defines the approach proposed to prevent ownership claim invalidation by means of additive attacks. Section 4 shows experimental results validating our proposal. Section 5 concludes this work.

2 Preliminaries

In this section we present part of the notation we will use throughout the paper, we give an overview of the basics of related watermarking techniques, and we discuss previous approaches proposed to deal with additive attacks.

2.1 Notation

According to Agrawal & Kiernan [2], let R be the relation to be marked, with: tuples r_j such that $j \in [0, \eta - 1]$, primary key PK, attributes a_i such that $i \in [0, \nu - 1]$, and scheme $R(PK, a_0, \ldots, a_{\nu-1})$. $r_j.a_i$ denotes the ith attribute of the jth tuple. η and ν are the number of tuples and the number of attributes in R respectively. ξ is the number of less significant bits (*lsb*) in the binary representation of an attribute value which can be marked. $\frac{1}{\gamma}$ is the Tuple Fraction (TF) which denotes the fraction of marked tuples, such that $\gamma \in [1, \eta]$. If the usability constraints are ignored, when $\gamma = 1$, all the tuples of the relation will be marked. ω is the number of marked tuples from the η tuples in R defined by the equation $\omega \approx \frac{\eta}{\gamma}$.

2.2 Background

The technique we propose in this paper is based on the image-based watermarking (IBW) approach for relational data presented in [7]. The latter mostly takes inspiration from two previous works: the one of Agrawal & Kiernan [2], and the one of Sardroudi & Ibrahim [13].

In 2002, Agrawal & Kiernan [2] defined the first relational data watermarking technique. Also called AHK algorithm, this approach embeds the marks in one of the ξ *lsb* of pseudo-randomly selected numeric attributes. In particular, once the attributes are determined, together with bit positions, and specific bit values, a meaningless bit pattern constituting the WM is embedded in R. The mark embedding locations depend on a secret key SK known only to the owner of the database. Also, the WM detection does not require either the access to the original data nor the WM, guaranteeing the technique's blindness. However, the AHK algorithm has been proven to be weakly resilient against subset attacks and data transformations. Moreover, the success of the detection phase may be penalized due to the meaningless of the watermarking information, and the data usability may be compromised as database constraints are ignored.

In [13], Sardroudi & Ibrahim defined a relational data watermarking scheme based on the AHK algorithm, that uses a binary image to generate the WM. The final reconstruction of the WM is done by performing a majority voting over each mark, which contributes to avoid the degradation of the WM that attacks based on data modification can cause. To make the scheme resilient against *subset reverse order attacks* [9], the pixels of the image used for WM generation, and the places to embed the marks in R, are chosen by using pseudo-random selection. Due to the pseudo-random nature of those processes, the embedding of the WM cannot be entirely achieved (even if all tuples of the relation are marked, which compromise data usability and make the WM perceptible, violating the imperceptibility requirement [5]).

Finally, as mentioned above, Gort et al., in [7], defined an IBW scheme close to the one presented by Sardroudi & Ibrahim, but able to overcome the limitations of the schemes presented in [13] and [2]. Indeed, Gort et al., increased the capacity of the WM (performing a controlled multi-attribute mark embedding, maintaining the quality of the data). Also, this scheme is proven to be robust against tuple deletion and addition attacks.

2.3 Main Approaches to deal with Additive Attacks

To deal with additive attacks, proposed techniques are mainly focused on two aspects: (i) taking advantage of the overlapping regions of the multiple WMs embedded in the database relation, or (ii) involving a Trusted Third Party (TTP) in the watermarking processes. Both approaches are based on scenarios that are hard to follow and can be easily compromised in practice. Below, the basics and limitations related to the approaches are given.

Overlapping regions of embedding. When an additive attack is performed, we can fall into one of the three following scenarios: (i) the attacker's WM entirely overwrites the owner's WM, (ii) some marks of both owner and attacker's WM have been embedded in the same positions (causing the overlapping of embedding regions), or (iii) the owner's WM and the attacker's WM do not collide at all, i.e., they are not embedded in same positions.

In the case in which the WMs do not collide, all ownership claims will be valid, annulling the process reliability. On the other hand, suspicion may raise if the attacker's WM entirely overwrites the owner's [1, 11]. Indeed, it is not usual that not even a single bit of the owner's WM being found in the data. Moreover, marks of different WMs occupying the same position may have the same value. Thus, an entirely WM overwriting changing all mark values is highly unlikely. Finally, when overlapping regions are present, the ownership claim competition is won by the one who inserted the last WM (i.e., the attacker) [1].

Consider the probability for embedding the marks in the same bits (c.f. Eq. (1) [1]), where, as previously mentioned, ω is the number of bits already marked by the data owner, and γ_A, ν_A, and ξ_A are the parameters used by the attacker to perform the additive attack. If the latter embedding parameters vary (as is expected, considering that if the attacker already knows the value of the parameters used by the data owner would not need to perform an additive attack), a low probability for embedding the marks in the same bits is expected. The more the probability gets closer to zero, the more the ownership assignment process gets more dubious, being even worse if some of the marks colluding present the same values.

$$P\{\text{success}|\omega\} = \left(1 - \frac{1}{2\gamma_A \nu_A \xi_A}\right)^{\omega} \tag{1}$$

Precisely, let A be a digital asset being protected by means of watermarking. The region allowed for the WM embedding in A is given by the function $\mathcal{Z}(\cdot)$, which returns an array of positions (the so-called *primary positions*). The notations W_O and W_A are used to refer to the WM embedded by the data owner and by the attacker respectively. The size of $\mathcal{Z}(A)$, W_O, and W_A can be obtained by using the function $n(\cdot)$. Figure 1 represents the scenarios given above, where the number of overlapping marks between W_O and W_A is given by δ.

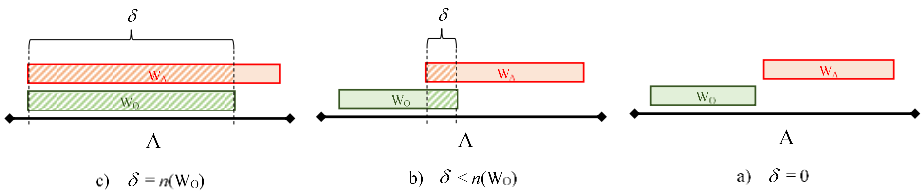


Fig. 1. Possible scenarios considering the overlapping between W_O and W_A.

Figure 1.a) is ruled by the probability of Eq. (1), which is expected to be low, or by the fact that $n(W_A) \approx n(\mathcal{Z}(A))$, which is unexpected if the attacker pretends to preserve the data usability. So, the complete overlapping of W_O by W_A, can be considered as a result of a successful brute force attack rather than by an additive attack. On the other hand, Figure 1.b) presents the case when some marks of W_O and W_A overlap. This scenario is mostly characterized by $n(\mathcal{Z}(A)) < n(W_O) + n(W_A)$. Also, under the previous condition, the probability of overlapping increases if $n(W_O) \approx n(W_A)$. Figure 1.c) corresponds to the case in which $n(\mathcal{Z}(A)) >> n(W_O) + n(W_A)$. The latter represents a critical situation since if both marks are embedded in A with no overlapping regions, there is no way to

determine which one was embedded first. Such situation cannot be avoided if the attacker uses a low size WM, even though, for the case of relational data it is not expected the attacker using a low size of W_A, since this would compromise its detection over time because of the degradation caused by benign updates. On the other hand, the data owner can successfully evade this situation by increasing the size of W_O as much as the usability of A tolerates.

Trusted Third Party involvement. Involving a TTP in the watermarking process means allowing a third person to assign the WM to be embedded, considering information from the data owner and adding other persons to the process (e.g., data buyers). Moreover, the TTP can be part of the generation of secret keys, among other important processes. Once the relation is watermarked, the TTP may also store copies of all the data involved [14].

Then, if another person wants to embed a WM on his/her data, comes to the TTP to perform the process. The TTP first checks if there is no other data owner already assigned to that data, and if it is not, proceeds to the WM embedding, secretly storing all data involved in the watermarking process once the task is concluded.

In this context, illegitimate owners may have no intention to present the data to the designated TTP for embedding their WM, or may claim the ownership of the data presenting their own WM to people unaware of the TTP existence. Moreover, involving a TTP is not always possible, can be quite expensive (it demands personal, time, technologies, and equipment) [11], and can lead to confidentiality concerns (e.g., in the case in which the TTP could have access to the data on its readable format). In the end, involving more people in the watermarking processes increases the probability of attacks.

2.4 Related Work

In 2003, Agrawal et al. [1] presented a deeper analysis of [2] in order to handle additive attacks in the AHK algorithm. They introduced Eq. (1) and showed how an attacker can manage to get a low number of overwritten bits with different mark values. Then, they considered both the idea of involving a TTP and of presenting the unwatermarked data, to solve false claims of ownership. Notice that the latter proposal can be easily compromised when the WM scheme can be inverted by creating a fake original data set and a fake WM [3].

In 2004, Li et al. [11] proposed to perform a WM embedding which aims to reach out into the maximum allowable distortion, thus reducing the possibility for the attacker to embed a second WM. This approach resulted to be vulnerable when $\xi_A \leq \xi$. Also, the attacker can always involve different parameters that allow his WM to be embedded without causing more distortion (e.g., by trying to preserve the attribute values distributions such as in [15]). On the other hand, Zhou et al. [17] presented an IBW technique where the WM to embed is generated from a binary image. This allows the generation of low aggressive WMs, and to embed a highly structural signal that can be

restored if attacks modifying the data are performed. The resilience of this technique to additive attacks is based on the involvement of a TTP.

In 2009, Gupta & Pieprzyk [8] defined a reversible watermarking technique, which allows obtaining the original data once the WM is extracted. The resilience of this technique to additive attacks is based on the involvement of a TTP. Notice that, in this case, once the WM is extracted the data will remain vulnerable to false ownership claims and other malicious operations. In 2010, Manjula & Settipalli [12] presented a technique that bases its resilience to additive attacks on tracking the overlapping marks. As previously mentioned, the success of this proposal will depend on the parameters used for the embedding of both WMs. Finally, in 2011, Hamadou et al. [10] presented a fragile technique that also bases its resilience to additive attacks on the involvement of a TTP.

3 The Extended Embedding Approach

In order to deal with false ownership claims by means of additive attacks, we exploit the WM overlapping regions (c.f. Figure 1.b)), and we define a non-colluded backup for the owner's marks by extending their embedding locations, determining the so-called secondary locations. In the case additive attacks are performed, the mark values stored in primary locations are corrected using the correspondent values recovered from secondary locations, making possible the identification of the WM.

3.1 Location linking structure

Figure 2 graphically shows the relation among the WM, the primary embedding locations, and the secondary ones. Each mark will be embedded multiple times on different primary locations $p_k^i: k \in [0, X_i - 1]$, being X_i the number of primary embedding for each mark. All primary locations corresponding to the same mark m_i, belonging to W_O, will be stored in the set $P_i: i \in [0, n(W_O) - 1]$. Linked to each primary location there is a set of secondary locations Sp_k^i, where each element is identified as $s_j: j \in [0, \ell_{k,i} - 1]$, being $\ell_{k,i}$ the number of secondary embeddings linked to the primary embedding k of the mark i.

Elements of secondary positions sets corresponding to different primary positions of the same mark can present elements in common (i.e., $Sp_a^i \cap Sp_b^i \geq \emptyset : a \neq b$), which enhances the possibility of properly restore the original mark value in the case in which it has been overwritten by an attacker. Eventually, the same secondary position can be assigned to different marks if they present the same value (i.e., if $(m_d = m_e) \rightarrow S_{P_d} \cap S_{P_e} \geq \emptyset : d \neq e$). On the other hand, the same secondary position can never be assigned to marks with different values, which will contradict the mark restoration even if no attacks are performed, compromising the WM synchronization and even its detection (i.e., if $(m_d \neq m_e) \rightarrow S_{P_d} \cap S_{P_e} = \emptyset : d \neq e$).

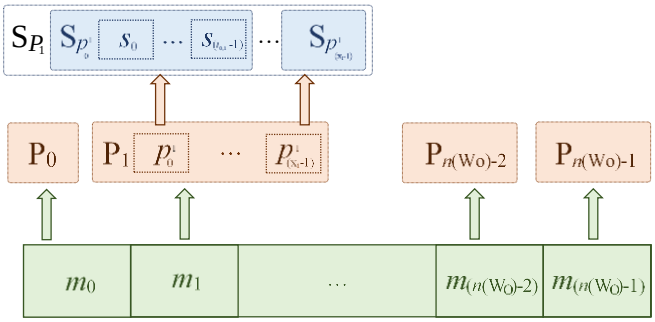


Fig. 2. Link between primary and secondary embedding locations.

3.2 Watermarking processes

The technique we propose is an extension of the conventional relational data watermarking technique in [7], performing an image-based WM generation and the embedding of the marks into the so-called primary locations. We propose, in this work, the module in charge of finding non-colluding locations for the secondary embedding, and the mechanism to embed the mark on those places.

Secondary locations depend on the virtual primary key of the tuple corresponding to the primary location (the virtual primary key *vpk* consist of a value generated to perform the WM synchronization involving the secrecy and privacy of the secret key SK and data identifying the tuple being analyzed, e.g., the relation's PK). This way, a strong link between the locations is created, avoiding the consequences of just increasing the embedding by changing the parameter values. The link among embedding locations allows higher control of the data usability during the WM embedding and improves the mark restoration effectiveness against additive attacks, compared to traditional approaches.

The starting point for secondary locations are those tuples satisfying the expression $vpk \ mod \ \gamma = 0$. Let us represent a generic tuple used for a first embedding as r_F. The ψth neighboring tuples to r_F (above and below of it) satisfying $vpk \ mod \ \gamma \neq 0$ (to avoid collusion with first locations) and $\varphi \neq -1$ will be considered for secondary embedding of the mark embedded in r_F. The symbol φ represents the variation of vpk with respect to its neighboring tuples. If the vpk constitutes a local minimum, then $\varphi = 0$ and the attributes considered for the mark embedding will be those below the mean of the numerical attributes of the tuple. For the case when vpk is a local maximum, then $\varphi = 1$ and the attributes considered for the embedding will be those above the mean of the numerical attributes of the tuple. The parameters controlling the collusion among locations in our approach are ψ and γ.

The WM extraction is performed similarly to the embedding but in the opposite direction (from the watermarked data to the reconstruction of the WM). The same parameter values are used and it is not necessary the original unwatermarked data nor the original source employed for the WM generation. Once a mark is extracted, the extraction of its copies stored on the correspondent secondary locations is performed.

Next, a majority voting is performed over the values extracted from the secondary locations and the primary mark. In case the values do not match, it is assumed that an additive attack was performed and the approach proceeds to the WM reconstruction.

4 Experimental Results

4.1 Experimental setup

We perform the experiments over the numeric relational dataset *Forest Cover Type* [16]. For the validation of the approach the first 30,000 tuples of the dataset were employed, as well as the 10 first attributes, to follow the methodology used in previous works and establishing fairly comparisons when the case demands. For the WM generation, the binary images shown in Table 1 were used.

Table 1. Images used as WM source.

Name	Sample	Size (pixels)
World Wildlife Fund (WWF)'s logo		40 x 45
Chinese character Dào's image		20 x 21

For measuring the differences between the embedded and extracted WMs is it employed the Correction Factor (CF) Eq.(2) where each pixel of the image employed to generate the embedded WM (given by Img_{org}) is compared to the ones of the image generated from the extracted WM (given by Img_{ext}). The symbols h and w represent the height and width of the images. The maximum value of CF is 100, which indicates the exact match of both images.

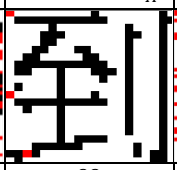
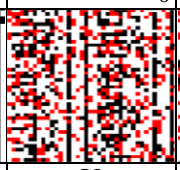
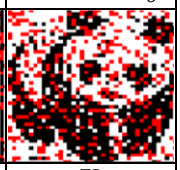
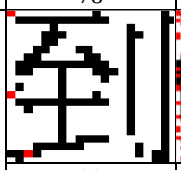
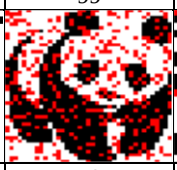
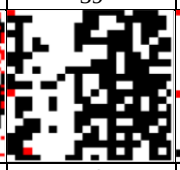
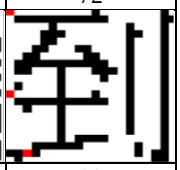
$$CF = \frac{\sum_{i=1}^{h}\sum_{j=1}^{w}(Img_{org}(i,j) \oplus \overline{Img_{ext}(i,j)})}{h \times w} \times 100 \quad (2)$$

4.2 Robustness against additive attack

Table 2 shows how by applying our approach the data owner's WM can be rebuilt from secondary embedding locations despite both watermarks being embedded over the same primary locations. In the table, *Embedded* W_O is the data owner's WM being embedded in the relation, *Embedded* W_A is the attacker's WM, *Unresilience* W_O constitutes the signal extracted by the watermarking technique with no secondary embedding locations, and *Resilience* W_O the WM recovered by applying our approach. For each case, the correspondent CF is also shown. The red pixels represent missed marks due to the partial embedding as a consequence of *pseudo-random selection*. The

experiment was performed changing the WMs belonging to both, the attacker and the data owner, to appreciate the role played by the WM's sizes.

Table 2. Images generated from the robustness experiments.

No.	*Embedded* W_O	*Embedded* W_A	*Unresilience* W_O	*Resilience* W_O
1				
	78	99	39	72
2				
	99	78	46	98

Finally, given that the complexity of our approach directly depends on the amount of data being protected, our scheme describes a performance proportional to the tuples of R, represented by $O(\eta)$.

5 Conclusion

In this paper, we proposed a watermarking technique for relational data based on secondary embedding locations to achieve resilience against additive attacks. Based on the analysis of the approaches proposed to deal with false ownership claims, we introduced a method that does not require involving a Trusted Third Party, avoiding the vulnerabilities and downsides of that type of solution. We were able to detect the presence of additive attacks and recover the owner's WM, gathering evidence to uncover the false claim of the attacker. As future work, we aim to analyze the relational watermarking technique we proposed in this paper with respect to *invertibility attacks* and extend it in order to completely prevent possible false claims of ownership.

Acknowledgements. This work was partially supported by the Ph.D. grant No. 714270 and the project grant No. PN 2017-01-7092 from CONACyT, Mexico.

References

1. Agrawal, R., Haas, P.J., Kiernan, J.: Watermarking Relational Data: Framework, Algorithms and Analysis. The VLDB Journal The International Journal on Very Large Data Bases 12(2), 157–169 (2003)

2. Agrawal, R., Kiernan, J.: Watermarking Relational Databases. In: VLDB'02: Proceedings of the 28th International Conference on Very Large Databases. pp. 155–166. Elsevier (2002)
3. Barni, M., Bartolini, F.: Watermarking systems engineering: enabling digital assets security and other applications. Crc Press (2004)
4. Berghel, H., O'Gorman, L.: Protecting ownership rights through digital watermarking. Computer 29(7), 101–103 (July 1996)
5. Cox, I., Miller, M., Bloom, J., Fridrich, J., Kalker, T.: Digital watermarking and steganography. Morgan kaufmann (2007)
6. Craver, S.A., Memon, N.D., Yeo, B.L., Yeung, M.M.: Can invisible watermarks resolve rightful ownerships? In: Storage and Retrieval for Image and Video Databases V. vol. 3022, pp. 310–321. International Society for Optics and Photonics (1997)
7. Gort, M.L.P., Uribe, C.F., Nummenmaa, J.: A Minimum Distortion: High Capacity Watermarking Technique for Relational Data. In: Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security, IH&MMSec 2017, Philadelphia, PA, USA, June 20-22, 2017. pp. 111–121 (2017)
8. Gupta, G., Pieprzyk, J.: Database relation watermarking resilient against secondary watermarking attacks. In: International Conference on Information Systems Security. pp. 222–236. Springer (2009)
9. Halder, R., Pal, S., Cortesi, A.: Watermarking techniques for relational databases: Survey, classification and comparison. J. UCS 16(21), 3164–3190 (2010)
10. Hamadou, A., Sun, X., Gao, L., Shah, S.A.: A fragile zero-watermarking technique for authentication of relational databases. International Journal of Digital Content Technology and its Applications 5(5) (2011)
11. Li, Y., Swarup, V., Jajodia, S.: Defending against additive attacks with maximal errors in watermarking relational databases. In: Research Directions in Data and Applications Security XVIII, pp. 81–94. Springer (2004)
12. Manjula, R., Settipalli, N.: A new relational watermarking scheme resilient to additive attacks. International Journal of Computer Applications 10(5), 1–7 (2010)
13. Sardroudi, H.M., Ibrahim, S.: A New Approach for Relational Database Watermarking Using Image. In: 5th International Conference on Computer Sciences and Convergence Information Technology. pp. 606–610 (2010)
14. Sencar, H.T., Memon, N.: Watermarking and ownership problem: a revisit. In: Proceedings of the 5th ACM workshop on Digital rights management. pp. 93–101. ACM (2005)
15. Sion, R., Atallah, M., Prabhakar, S.: Rights protection for relational data. IEEE transactions on knowledge and data engineering 16(12), 1509–1525 (2004)
16. University., C.S.: Forest covertype, the uci kdd archive. (Jun 1999). https://doi.org/Information and Computer Science. University of California, Irvine, http://kdd.ics.uci.edu/databases/covertype/covertype.html
17. Zhou, X., Huang, M., Peng, Z.: An additive-attack-proof watermarking mechanism for databases' copyrights protection using image. In: Proceedings of the 2007 ACM symposium on Applied computing. pp. 254–258. ACM (2007)