数据库水印模型及其算法综述

王春东[1,3],杨 蕾[1,3],万福金[2],刘哲理[2],朱立坤[1,3]
(1. 天津理工大学计算机科学与工程学院,天津 300384; 2. 南开大学网络空间安全学院,天津 300350;
3. 天津理工大学"计算机视觉与系统"教育部重点实验室,天津 300384)

摘 要: 大数据时代迅速发展的同时,如何解决数据的版权保护、信息保存、叛逆溯源等安全问题,也变得越来越重要.数据库水印技术是最有效的防护技术,其面临的最大挑战是如何在不影响用户使用的前提下嵌入水印,且保证在不影响水印信息提取的前提下抵挡各种常见的水印攻击.本文首先描述了数据库水印的基本概念和工作原理;然后,介绍了常见的数据库水印技术,并进行了分类;进而,总结了针对数据库水印的攻击行为,分析了特定水印技术的抗攻击性;接着详细分析了基于失真与基于无失真的典型水印技术,并简要分析了数据库指纹的现有技术手段;最后,指出各种现有的技术研究的关键与不足,提出了下一步可研究的方向.

关键词: 数据库水印; 数据库指纹; 版权声明; 鲁棒性
中图分类号: TP309.2 **文献标识码:** A **文章编号:** 0372-2112 (2019)04-0946-09
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.04.022

Survey on Database Watermarking Models and Algorithms

WANG Chun-dong[1,3], YANG Lei[1,3], WAN Fu-jin[2], LIU Zhe-li[2], ZHU Li-kun[1,3]
(1. *School of Computer Science and Engineering, Tianjin University of Technology, Tianjin* 300384, *China*;
2. *College of Cyberspace Security, Nankai University, Tianjin* 300350, *China*;
3. *Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin University of Technology, Tianjin* 300384, *China*)

Abstract: With the rapid development of big data, it becomes more and more important to effectively solve the problems of copyright protection of data, preservation of data information, traceability of data, etc. Database watermarking is the most effective protection technique. The biggest challenge of this technology is how to carry out the embedded watermarking without affecting the data, and how to withstand the various attacks without affecting the extraction of watermark information. In this paper, we first describe the basic concepts and principles of database watermarking; Then, we introduce and classify common database watermarking techniques; Next, we summarize the attack behavior against database watermarking and analyze the anti-attack ability of specific watermarking technology; Moreover, we analyze the typical watermarking technologies based on distortion or not and the existing technical methods of database fingerprint briefly; Finally, we point out the key and shortcomings of current researches, and propose the development direction in the future.

Key words: database watermarking; database fingerprint; copyright statement; robustness

1 引言

互联网技术的迅速发展,云计算、物联网、人工智能等数字信息技术应运而生,这个过程会随之产生大量数据,因此对数据库服务和存储等系统的要求越演愈烈.这些数据会发布在公共平台,会被分发给下属用户,会在网络平台上被访问以及用于其他目的,可是同时也给数据安全问题也带来了巨大的挑战.攻击者或者盗版者很容易对数据的所有权进行伪造,泄密者会改变数据源从而逃脱泄密数据的责任.为了解决这类安全问题,数据库水印技术应运而生.该技术可以有效解决侵权与篡改信息问题,而后期从数据库水印技术延伸而来的数据库指纹技术则可以有效地进行数据溯源.因此受到各界越来越多的关注.

水印是一种被分发者嵌入到数据库信息中以此来进行版权追踪的秘密信息,这种信息可以是图像、符号、

收稿日期:2018-05-28;修回日期:2018-09-09;责任编辑:孙瑶
基金项目:国家自然科学基金委通用技术基础研究基金(No. U1536122);天津市科委重大项目(No. 15ZXDSGX00030);国家自然科学基金(No. 61672300);天津市自然科学基金(No. 16JCYBJC15500);教育部和中国移动科学研究基金(No. MCM20170403)

文字等.通常情况下,这种信息是不对外显示的,在攻击者发动攻击时该信息可以基本上不被攻击破坏.在有人恶意对其进行版权侵犯时,数据持有方可以通过一系列的提取算法完整的提取出数据库中嵌入的原始水印,以此来宣告自己对该数据的所有权.之前的数字水印技术的研究主要集中在图像和音频等多媒体水印方面.近年来随着大数据的日益盛行,关系数据库中水印技术的研究显得愈加重要.

2000 年,Khanna S 等率先提出了利用数字水印技术对数据库进行安全控制的新思路,使得数据库水印技术引起了研究者的关注. Agrawal R[1] 在 2002 年提出了第一个数据库水印算法,主要通过修改数值型数据的最低有效位来进行水印的嵌入,允许数据产生少量的失真.在此之后,诸多数据库水印算法被陆续提出[2~60].初期的数据库水印算法只是为了对数据库进行简单的版权证明,较少考虑各种复杂的数据库水印攻击. Zhang[31] 等人在 2006 年首次提出可逆的数据库水印方案,使得原始数据库可以再嵌入水印以后依旧可以被还原;2007 年 Li [40] 和 Tsai[44] 等人陆续提出基于属性标记以及基于哈希值等通过原始数据库产生待嵌入水印的数据库水印算法,逐渐开始进行无失真水印技术的研究,可以在最大程度上保证演示数据的完整性,算法的抗攻击能力也随之增强;早在 2003 年,数据库水印模型就已经开始被应用于数据库指纹技术,技术目的从最初的版权证明演变为盗版追踪和溯源,后期一些基于块的数据库指纹技术也陆续围绕此提出.

2　基本定义与概念

2.1　相关定义

数据库指纹[6] 是从数据库水印分支中得来的,数据库水印和数据库指纹这两大技术为数据库的隐私保护起到了关键作用.下面综合各种文献,给出了与数据库水印技术相关的一些描述性定义.

定义 1(关系数据库)　设关系数据库为 $D(P,A_1,\cdots,A_i,\cdots,A_n)$,其中:$P$ 为主键;$A_1,\cdots,A_i,\cdots,A_n$ 为 n 个属性列;D 由 m 个元组 $r_2,\cdots,r_j,\cdots,r_m$ 组成,每个元组 r 都存在主键 $r.P$ 和 n 个属性值 $r.A_2,\cdots,r.A_i,\cdots,r.A_n$,可以得到 $m\times n$ 个属性值.

定义 2(数据库水印)　数据库水印技术是将水印信息嵌入到数据库.它比普通嵌入到多媒体中的水印要求要高一些,除了需要在提取阶段可以基本上不受损耗的被发布者提取出来,还需要在嵌入数据库的过程中尽可能小的修改原始数据的使用价值,确保信息可以被用户正常使用.

定义 3(数据库指纹)　数据库指纹是从数据库水印技术延伸而来,它是将不同的标志性的信息——指纹,利用数据库水印的技术嵌入到数据库中,然后将嵌入不同指纹信息的数据库依次分发给各个用户 A、B 等人.在市场中发现其他盗版数据库时,可以根据数据库中的指纹信息来确定盗版数据库的来源,从而达到保护版权的作用.

2.2　特点

一般来讲,数据库水印具有以下的特点[4~11].

(1)隐蔽性　为了防止嵌入的水印会影响数据使用者的使用,应该确保数据在嵌入水印以后对原数据使用程度的影响可以忽略不计,确保水印对于使用者是隐蔽和不可见的.

(2)安全性　水印的主要目的就是来证明数据库中数据的所有权,因此要确保被嵌入的水印很难被伪造或者修改.而且攻击者在不知道秘密参数的情况下难以完整有效的确定嵌入水印的位置并提取出水印信息.

(3)可检测性　水印信息除了不被破坏以外,还需要保证其可以通过一定的提取算法被完整的提取出来,以此来宣告数据拥有方对其的版权.

(4)鲁棒性　数据库中的数据受到攻击方的恶意攻击后,数据持有方依旧可以完整的提取出水印信息,以此验证其版权.

3　分类

3.1　分类

截止到目前,研究主要围绕数据库水印模型和数据库指纹算法展开,其中,数据库水印模型通过比较其是否对底层数据引入了失真,可以将其分为基于失真、基于无失真两类[12,13],如表 1 所示.

基于失真的数据库水印技术一般是利用字符串、图像、符号等信息作为嵌入水印信息,而基于无失真的数据库水印技术则一般是利用原始数据库的信息生成水印信息.数据库指纹技术则一般是对不同的用户嵌入不同的有意义的字符串.

3.2　适用范围

目前对数据库水印技术的研究只是针对特定应用场景提出特定的技术方案,而不同的数字水印技术方案的适用范围是不同的,表 2 总结了现有的各类技术的水印信息和其适用范围[14~20].通过以上表 2 对典型方案的分类与比较,可以发现:

(1)基于失真的数据库水印方案的水印信息种类较多,包括字符串、图像等;基于无失真的数据库水印方案则一般是利用原始数据库本身生成水印信息或者添加伪行伪列等方法.

(2)基于失真的数据库水印方案所针对的数据库

信息的数据类型一般是数值型或者分类数据等特定的数据类型;基于无失真的数据库水印方案一般针对的数据类型可以是数值型,也可以是非数值型,使用范围较为广泛. 而数据库指纹技术则主要应用于数值型数据.

(3)由于现有数据库水印技术只适用于特定的应用,缺少通用的数据库水印模型,所以在实际应用中需要综合各类水印技术来构建满足实际应用的水印模型.

表 1 典型方案分类与比较

classification	technology	Typical Schemes
Database watermarking technology based on distortion	based on Meaningless string	[Agrawal,2002][1]; [Agrawal,2003a][2]; [Agrawal,2003b][3];
	based on image	[Zhang,2004][21]; [Chen,2008][23]; [Wang,2008a][24]; [Wang,2008b][25]; [Hu,2009][26]; [Zhou et al,2007][27];
	based on data of classification	[Sion,2005][28]; [Sion,2004][29];
	Based on reversible	[Zhang,2006][31]; [Farfoura,2012][32]; [Chang,2012][33]; [Gupta,2007][34];
	Based on Availability constraint model	[Khurram,2013][35]; [Kamran,2013][36]; [Kamran,2012][37,38]; [Kamran,2015][39];
Watermarking technique based on undistorted database	based on Hash	[Li,2004][40]; [Bhattacharya,2009a][41]; [Bhattacharya,2009b][42]; [Li,2007][43];
	Based on R tree arrangement	[Kamel,2009][50];
	Zero watermarking database	[Khan,2013][51]; [Le,2016][52];
	Based on pseudo tuple	[Pournaghshband,2008][53];
	Based on virtual attributes	[Prasannakumari,2009][54];
	Block oriented database fingerprinting	[Liu,2004][60];

表 2 现有方案应用场景

technology	numeric type	Non numerical type
Database watermarking technology based on meaningless string[1~3]	✓	
Watermarking technology based on image database[21,23~27]	✓	✓
Database watermarking technology based on Classification[28,29]	✓	
Reversible database watermarking technology[31~34]	✓	
Database watermarking technology based on availability constraints[35~39]	✓	✓
Database watermarking technology based on hash value[40~43]	✓	✓
Database watermarking technology based on attribute label[44]	✓	✓
Watermarking technology based on binary database[46~49]	✓	
Database watermarking technology based on R tree[50]	✓	
Zero watermarking database watermarking technology[51,52]	✓	
Watermarking technology of database based on pseudo tuple[53]	✓	✓
Database watermarking technology based on virtual attributes[54]	✓	✓
Database fingerprint technology[55~57]	✓	

4 攻击

数据库水印易于遭受攻击,使得水印遭受破坏或者丢失,因此在大多数水印在设计过程中会考虑抵御某些已存在的水印攻击. 本节总结了各类常见的数据库水印攻击以及现有技术的抗攻击现状.

4.1 常见攻击

数据库水印攻击的目的是破坏原始嵌入水印信息,或将原始水印信息更换为携带攻击方版权信息的水印. 手段主要包括元组攻击、属性攻击、虚拟水印攻击和其他攻击等.

元组攻击是数据库水印攻击中针对元组进行的攻击. 该攻击主要包括元组选择攻击、添加攻击以及修改攻击三类.

属性攻击是数据库水印攻击中针对属性进行的攻

击. 该攻击主要包括属性更改攻击、删除攻击、添加攻击以及重组攻击四类.

虚拟水印是指攻击者通过添加虚假的水印来宣称自己对该数据库的版权所有. 虚拟水印攻击分为可逆性攻击和重复水印攻击两类.

除此以外,常见攻击还有混合匹配攻击、暴力破解攻击和多方攻击等.

4.2　抗攻击现状

表 3 为现有技术可抵挡常见攻击类型,从上表 3 可以看出:

(1)大部分水印方案都可以抵挡删除攻击、修改攻击和插入攻击等常见攻击类型.

(2)基于二进制的水印方案和基于 R 树的水印方案分别可以有效抵挡移位攻击和更新攻击;伪行和伪列可能会对特定的攻击类型有较强的抵抗能力;数据库指纹技术可以抵挡较多类型的攻击.

表 3　现有技术可抵挡常见攻击类型

Technology	a	b	c	d	e	f	g	h	i	j
Meaningless string	√	√	√	√						
Image					√	√				
Classification			√				√	√		
Reversible technology			√					√	√	
Availability constraints			√				√	√		√
Hash value			√				√	√		
Attribute label							√	√		
Binary database			√				√	√		
R tree			√				√	√		
Zero watermark			√				√			
Pseudo tuple			√							
Virtual attributes			√				√			
Database fingerprint		√		√			√	√		

③每个水印方案都是针对特定的应用场景,因此没有任何一个方案可以抵御大多数攻击. 实际应用中一般利用多种水印算法的结合才能够最大程度的抵挡各种攻击. (Hybrid matching attack——a;Bit flipping attack——b;Add attack——c;Reversible attack——d;Vertical data attack——e;Attribute recombination——f;Delete attack——g;Modification attack——h;Tuple selection attack——i;Hybrid attack——j;Shift attack——k)

5　基于失真的数据库水印算法

对基于失真的数据库水印技术中基于无意义字符串、基于图像、基于分类数据、基于可逆、基于可用性约束等水印方案进行详细的介绍和分析.

5.1　基于无意义字符串的水印算法

数据库水印算法是在 2002 年由 Agrawal R 和 Kiernan J[1,2,3] 首次提出的. 它是对基于多媒体数字水印的改进,只标记数值型的数据. 表 4 是实施算法时所需要的重要参数.

表 4　算法中的重要参数

变量	含义
η	数据库中总元组数
ν	数据库中可以被标记的元组数
ξ	可以被标记的最低属性位
$1/Y$	被标记元组百分比
ω	被标记元组数目
α	水印检测阶段的重要性

在水印嵌入过程中,利用 $F(r.P)$ 选择标记的元组和属性,对于关系 R 中的每个元组 r,若 $F(r.P) \bmod Y=0$,标记元组 r;计算 $i=F(r.P) \bmod \nu$,标记属性 A_i;计算 $j=F(r.P) \bmod \xi$,标记该属性第 j 位;从而确定出需要标记的具体位置. (其中 $F(r.P)=H(K^{\circ}H(K^{\circ}r.P))$,$^{\circ}$表示连接符号). 对主键 p 和私有密钥 K 连接以后进行哈希映射. 如果该哈希值为偶数,则第 j 位置为 0,如果该哈希值为奇数,则第 j 位置为 1,按照该算法将关系数据库中待标记的属性的比特位置全部置为 0 或者 1,作为标记位. 在这些需要标记的元组都进行标记以后,这些标记比特位的分布就是版权标记信息.

在水印检测过程中,若 $F(s.P) \bmod Y=0$,元组 s 被标记;计算 $i=F(s.P) \bmod v$,属性 A_i 被标记;计算 $j=F(s.P) \bmod \xi$,属性 A_i 第 j 位被标记. 确定出哪些位是标记位. 根据对元组的主键和密钥的连接值进行哈希计算出嵌入的水印值,如果 S 中的该位置与计算出的值相等,检测出的 *matchcount* 加 1,都计算完毕后,查看 *matchcount* 是否超过门限值 τ,若超过则证明该怀疑的数据库被盗用. 在进行水印的提取时需要知道嵌入水印的比例和密钥等信息,最后使用多数选举法进行提取. 该算法可以有效地抵挡混合匹配攻击、位翻转攻击、添加攻击和可逆性攻击等常见攻击. Sion R 等[5] 利用秘密排序对 AHK 算法进行了进一步的研究. 秦等人利用混沌方程对之前算法中水印的生成阶段进行了改进[19,20]. 该算法可以有效地抵挡混合匹配攻击、位翻转攻击、添加攻击和可逆性攻击等常见攻击.

上述基于无意义字符串的水印算法针对的数值型数据,通过改变浮点型数据的冗余位来嵌入信息. 但是,这种算法还是会在一定程度上影响原始数据. 在实际应用中需要尽可能完整的保留原始数据,这时需要的是基于无失真的水印算法.

5.2 基于图像信息的水印算法

根据所嵌入水印信息的类型不同,水印信息也从最初的字符串演变成图像[21~25]. 王等[24,25]就描述了一种基于图像的数据库水印算法. Hu 等[26]将原始的水印图像转化为一定长度的位流(EMC,加密标记代码),接着按照 Wang 等人的嵌入算法进行水印的嵌入. 一种使用 BMP 格式嵌入图像的水印方案由 Zhou[27] 在 2007 提出.

5.3 基于分类数据的水印算法

分类数据[28,29]是数据库中的一种数据类型,它是反映事物类别的数据类型,具有有限个无序的值,因此是离散数据. 比如:属性——城市,有天津、北京、上海等值,而基于分类数据的数据库水印是将水印隐藏到分类数据中,在这种方案下,对于技术的要求不再是数值型尽量减少对数据的改动,而是尽量减少改动的数据项.

5.4 基于可逆的水印算法

对于可逆数据库水印的研究主要包括基于直方图扩展[31]的方法,基于时间戳[32]的方法,基于 SVR[33]的方法等. Gupta 等人[34]提出一种可逆水印方案,该算法具有较强的鲁棒性. 能有效抵挡删除攻击,添加攻击和修改攻击. 基于 DEW(差分扩展)[35]的扩展水印技术是对之前水印技术的一种改进,该技术中利用水印关系的均值和标准偏差来评估属性中的失真程度.

基于可逆的数据库水印算法在提取水印信息的同时可以恢复原始数据,但是算法的复杂度较高,效率较低,很难满足当今大数据下的数据溯源要求.

6 基于无失真的数据库水印算法

基于无失真的数据库水印技术中主要包括基于元组哈希值、基于属性标记、基于二进制字符串、基于 R 树、基于零水印以及伪行伪列等水印方案.

6.1 基于哈希值的水印算法

在 Li[40]的方案中,元组的划分是基于使用主键和密钥参数化的哈希值,而后来的研究[41,42]中,分区则是基于分类属性值,分区后,计算每个组的元组级别和元组的哈希值. 此外,Li[43]在 2007 年建议使用基于 Myrvold 和 Ruskeys 线性排列未排序算法执行元组的位置交换,以此来增加嵌入能力.

6.2 基于 R 树的水印算法

相对于传统水印方案,Kamel[50]提出来一种基于 R 树的水印技术,这种方案利用了 R 树不将条件置于节点内条目的顺序的事实. 将 R 树节点中的条目所有可能的排列顺序与水印的所有可能值一一映射. 水印被假定为数值,之前提出的映射使用具有阶乘值的可变基数的编号系统. 恶意攻击的检测率一般取决于数据的分布以及 R 树节点的大小.

6.3 零水印算法

后来 Khan 等[51]学者利用了零水印技术来进行版权的验证,相对之前传统数据库水印模型利用数据库中属性值的 LSB 来进行水印的嵌入,有些类似于身份证号码这些数据对数据精度的要求相对较高,基本上不能存在失真,因此对于这些数据的数据库可以采用零水印技术. 零水印算法除了上述用于版权验证,还能保护数据库中面部图像的隐私[52].

6.4 基于伪元组的水印算法

为了在嵌入水印的过程中不影响原始的数据信息,可以利用生成伪元组[53]并将其错误的插入到数据库中来嵌入水印. 该种水印的生成需要数据拥有方的一些信息. 伪元组的创建是利用对非候选属性和候选属性的敏感度级别进行处理生成的.

这种方案可以有效的抵挡添加元组和通过添加自己的伪元组进行虚假所有权证明等基本攻击.

6.5 基于虚拟属性的水印算法

后来的学者[54]提出了另一种基于属性的数据库水印技术,该方案的水印模型如图 1 所示,其核心是通过在关系中插入虚拟属性 V_A. 该虚拟属性是利用特定分区的聚合函数生成的,而且会将其它属性的奇偶校验和该聚合函数进行连接作为水印. 原始数据关系按照一定的规则进行分区后,获得 n 个非重叠区域,分别对这些分区独立地执行虚拟属性插入. 虽然这种水印方案是脆弱的,并且可以容易地检测到任何删除、插入或改变攻击,但是它无法有效抵挡水印删除攻击.

上述基于无失真的数据库水印技术研究可以在最大程度上保证原始数据的可用性,但是,这些算法的适用数据类型的局限性依旧存在. 因此,需要一种新的技术,在保证数据可用性的同时,突破对数据类型的限制,并且可以抵御常见的数据库攻击类型.

7 数据库指纹技术

数据库指纹与数据库水印两种技术的侧重点不尽相同,数据库水印技术发展较早,主要目的是为了买家在发生版权纠纷时验证其对数据库的版权所有;而数据库指纹则是为了用标记识别分发的叛徒,对其进行盗版溯源.

Liu 等[57]提出了一种面向块的数据库指纹水印算法. 该算法主要将指纹为嵌入到二维图像中,例如:假设主键为 P,如表 5 所示,首先提取出最后三位将其重新组合为表 6,此时表 7 即为 2×2 的指纹块.

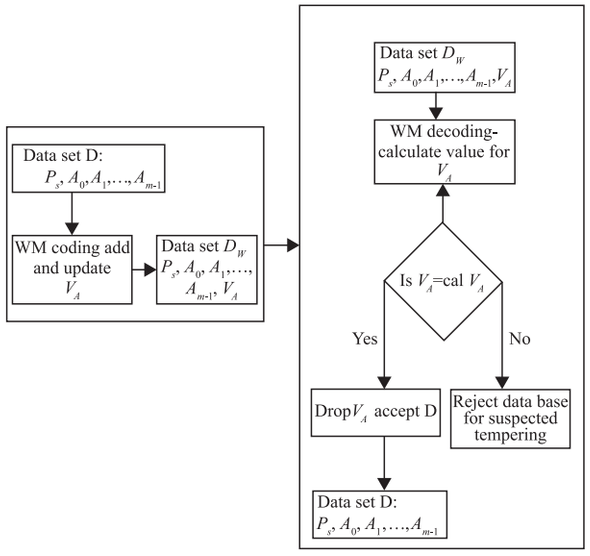

图1　水印系统模型框架

表 5　关系数据部分

P	A_1	A_2
1	01100011	00001001
2	10000010	00100111
3	01001111	10010001
4	00000000	00000101

表 6　可利用指纹位

011001
010111
111001
000101

表 7　6 个 2×2 的块

01 01	10 01	01 11
11 00	10 01	01 01

在指纹嵌入过程中,首先为每个买家生成指纹,利用伪随机序列发生器选择一个阈值 r. 接着将数据库的属性为划分为大小为 2×2 的块. 根据哈希值与 β 的模运算结果确定指纹嵌入位置,接着对于每个块分别进行指纹的嵌入. 经验证,该方案可以有效的抵挡子集添加、子集删除、修改元组和属性等常见攻击.

8　研究不足与未来趋势

在大数据时代,由于数据库水印技术[58]对于数据库的安全的必要性,有关于这个领域的研究[59~61]和应用显得尤为重要. 近十几年来,虽然国内外有很多有关于数据库水印技术的研究,但是仍然有很多值得深入探索的问题. 本节针对前文中现有的典型方案进行分析总结以后,提出了现有方案的不足.

(1)数据类型. 现有技术中所研究的数据类型大多数是针对数值型数据,利用改变数据的冗余位来嵌入水印信息;少部分技术是针对文本型数据,利用不可见字符来嵌入水印. 缺少一种算法可以突破数据类型的限制.

(2)冗余空间. 数据库水印技术中嵌入水印的关键就是寻找数据的冗余空间,目前的数据库水印技术大多数在数据本身寻找冗余空间,嵌入水印效率较低.

(3)数据失真. 大多数水印技术极其容易影响数据的可用性和真实性.

(4)鲁棒性. 每种技术都难以有效地抵挡所有常见的数据库攻击类型.

(5)时间和成本. 目前的水印技术通常算法计算量较大,时间和成本代价较高.

基于上述现有技术的不足之处,可以看出,数据库水印技术接下来的研究方向主要表现在以下几个方面.

(1)突破数据类型限制. 实际应用中,数据的类型往往是复杂多样的. 因此,需要一种通用有效的数据库水印算法,突破传统数据库水印算法中对数据类型的限制.

(2)寻找数据库中的冗余空间. 为了保证在嵌入水印过程中的可以有效地衡量载体数据的失真程度,需要准确的数学模型和描述可以量化出数据库的冗余空间.

(3)保证数据失真最小. 可以寻找一种特殊的方式,将水印信息隐藏在数据之间的关系中,这样可以保证数据不被破坏.

(4)增强数据库水印的鲁棒性. 针对水印生成阶段中的水印预处理阶段,将传统的水印生产方案结合混沌序列和聚类等方式对水印进行预处理,可以使得水印预处理阶段的安全性更高.

(5)降低成本和时间. 之前的大部分数据库水印算法都主要考虑其鲁棒性,却极少考虑实验的时间. 可是当今大数据时代的数据量越来越大,消耗过多时间技术的价值显然不高. 因此也要考虑算法的时间和成本.

9　总结

本文主要介绍了关系数据库中各种类型的水印与指纹技术. 主要利用对底层数据是否引入失真对其技术进行了分类,大多数基于失真的水印技术的主要目的是为了保护所有权,而基于无失真的水印技术则是脆弱的,主要目的是为了保护数据库信息的完整性. 着重分析了几种典型水印方案的水印生成方式、嵌入方

式,比较了每种类型的方案所能应对的攻击和应用场景.对未来数据库水印的发展方向做出了一定的预测和展望.

参考文献

[1] AGRAWAL R, KIERNAN J. Watermarking relational databases[A]. Proceedings of the 28th International Conference on Very Large Data Bases (VLDB '02)[C]. Hong Kong, China: VLDB Endowment, 2002. 155 – 166.

[2] AGRAWAL R, HAAS P, KIERNAN J. A system for watermarking relational databases[A]. Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data (SIGMOD'03)[C]. San Diego, California: ACM Press, 2003. 674 – 674.

[3] AGRAWAL R, HAAS P J, KIERNAN J. Watermarking relational data: framework, algorithms and analysis[J]. The VLDB Journal, 2003, 12(2): 157 – 169.

[4] SHEHAB M, BERTINO E, GHAFOOR A. Watermarking relational databases using optimization-based techniques [J]. IEEE Transactions on Knowledge and Data Engineering, 2008, 20(1): 116 – 129.

[5] SION R, ATALLAH Mikhail, PRABHAKAR Sunil. Rights protection for relational date[J]. IEEE Transactions on Knowledge and Data Engineering, 2004, 16(6): 1509 – 1525.

[6] MILANO D. Content control: Digital watermarking and fingerprinting[A]. White Paper, Rhozet, a Business Unit of Harmonic Inc[P/OL]. http://www.rhozet.com/whitepapers/Fingerprinting Watermarking.pdf, Last Accessed, 2012 – 05 – 30.

[7] GUO H, LI Y, LIU A, et al. A fragile watermarking scheme for detecting malicious modifications of database relations [J]. Information Sciences, 2006, 176(10): 1350 – 1378.

[8] KSHATRIYA M S, SANE S S. A study of watermarking relational databases[J]. International Journal of Application or Innovation in Engineering & Management (IJAIEM), 2014, 3(10): 154 – 158.

[9] MEHTA B B, ASWAR H D. Watermarking for security in database: A review[A]. Conference on IT in Business, Industry and Government (CSIBIG)[C]. US: IEEE, 2014. 1 – 6.

[10] HALDER R, PAL S, CORTESI A. Watermarking techniques for relational databases: Survey, classification and comparison[J]. Journal of UCS, 2010, 16(21): 3164 – 3190.

[11] DWIVEDI A K, SHARMA B K, VYAS A K. Watermarking techniques for ownership protection of relational databases[J]. International Journal of Emerging Technology and Advanced Engineering, 2014, 4(1): 368 – 375.

[12] JOSE N, UPPAR A S. A Secured Approach for Numerical Relational Databases Using Robust and Reversible Watermarking Technique[J]. International Journal of Emerging Research in Management & Technology, 2016, 5(5): 223 – 227.

[13] HAN A X, CHEN Y. Research on database watermarking algorithm based on genetic algorithm[A]. Applied Mechanics and Materials[C]. US: Trans Tech Publications, 2014, 687: 1297 – 1299.

[14] BHESANIYA M H, THANKI K. Watermarking of relational databases[J]. International Journal for Research in Technological Studies, 2013, 1(1): 11 – 16.

[15] KAMRAN M, SUHAIL S, FAROOQ M. A robust, distortion minimizing technique for watermarking relational databases using once-for-all usability constraints[J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(12): 2694 – 2707.

[16] XIE M R, WU C C, SHEN J J, et al. A survey of data distortion watermarking relational databases[J]. International Journal of Network Security, 2016, 18(6): 1022 – 1033.

[17] MOHANPURKAR A A, JOSHI M S. Applying watermarking for copyright protection, traitor identification and joint ownership: A review[A]. World Congress on Information and Communication Technologies (WICT)[C]. US: IEEE, 2011. 1014 – 1019.

[18] BHESANIYA M H, RATHOD J, THANKI K. Various approaches for watermarking of relational database[J]. International Journal of Engineering Science and Innovative Technology, 2014, 3(1): 215 – 220.

[19] QIN Z, YING Y, JIA-JIN L E, et al. Watermark based copyright protection of outsourced database[A]. The 10th International Database Engineering and Applications Symposium (IDEAS06)[C]. US: IEEE, 2006. 301 – 308.

[20] MENGMENG F, JIANMIN L, ZHENRONG L. A new relational database watermarking algorithm[J]. Computer Applications and Software, 2015, 6: 62.

[21] ZHANG Z H, JIN X M, WANG J M, et al. Watermarking relational database using image[A]. Proceedings of 2004 International Conference on Machine Learning and Cybernetics[C]. US: IEEE, 2004. 1739 – 1744.

[22] SARDROUDI H M, IBRAHIM S. A new approach for relational database watermarking using image[A]. The 5th International Conference on Computer Sciences and Convergence Information Technology (ICCIT)[C]. US: IEEE, 2010. 606 – 610.

[23] CHEN X, CHEN P, HE Y, et al. A self-resilience digital image watermark based on relational database[A]. International Symposium on Knowledge Acquisition and Modeling (KAM08)[C]. US: IEEE, 2008. 698 – 702.

[24] WANG C, WANG J, ZHOU M, et al. Atbam: An arnold transform based method on watermarking relational data [A]. International Conference on Multimedia and Ubiquitous Engineering (MUE'08) [C]. US: IEEE, 2008. 263 – 270.

[25] WANG H, CUI X, CAO Z. A speech based algorithm for watermarking relational databases [A]. International Symposiums on Information Processing (ISIP) [C]. US: IEEE, 2008. 603 – 606.

[26] HU Z, CAO Z, SUN J. An image based algorithm for watermarking relational databases [A]. International Conference on Measuring Technology and Mechatronics Automation (ICMTMA'09) [C]. US: IEEE, 2009. 425 – 428.

[27] ZHOU X, HUANG M, PENG Z. An additive-attack-proof watermarking mechanism for databases' copyrights protection using image [A]. Proceedings of the 2007 ACM Symposium on Applied Computing [C]. US: ACM, 2007. 254 – 258.

[28] SION R. Proving ownership over categorical data [A]. Proceedings of the 20th International Conference on Data Engineering [C]. US: IEEE, 2004. 584 – 595.

[29] SION R, ATALLAH M, PRABHAKAR S. Rights protection for categorical data [J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(7): 912 – 926.

[30] KUMAR M, VERMA O P, SAXENA A. Elliptic curve cryptography (ECC) based relational database watermarking [J]. International Journal of Computer Applications, 2016, 154(6): 20 – 28.

[31] ZHANG Y, YANG B, NIU X M. Reversible watermarking for relational database authentication [J]. Journal of Computers, 2006, 17(2): 59 – 66.

[32] FARFOURA M E, HORNG S J, LAI J L, et al. A blind reversible method for watermarking relational databases based on a time-stamping protocol [J]. Expert Systems with Applications, 2012, 39(3): 3185 – 3196.

[33] CHANG J N, WU H C. Reversible fragile database watermarking technology using difference expansion based on svr prediction [A]. International Symposium on Computer, Consumer and Control (IS3C) [C]. US: IEEE, 2012. 690 – 693.

[34] GUPTA G, PIEPRZYK J. Database relation watermarking resilient against secondary watermarking attacks [A]. International Conference on Information Systems Security [C]. Berlin Heidelberg: Springer, 2009. 222 – 236.

[35] JAWAD K, KHAN A. Genetic algorithm and difference expansion based reversible watermarking for relational databases [J]. Journal of Systems and Software, 2013, 86 (11): 2742 – 2753.

[36] KAMRAN M, SUHAIL S, FAROOQ M. A robust, distortion minimizing technique for watermarking relational databases using once-for-all usability constraints [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 25(12): 2694 – 2707.

[37] KAMRAN M, FAROOQ M. An information-preserving watermarking scheme for right protection of EMR systems [J]. IEEE Transactions on Knowledge and Data Engineering, 2012, 24(11): 1950 – 1962.

[38] KAMRAN M, FAROOQ M. A formal usability constraints model for watermarking of outsourced datasets [J]. IEEE Transactions on Information Forensics and Security, 2013, 8(6): 1061 – 1072.

[39] IFTIKHAR S, KAMRAN M, ANWAR Z. RRW—a robust and reversible watermarking technique for relational data [J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(4): 1132 – 1145.

[40] LI Y, GUO H, JAJODIA S. Tamper detection and localization for categorical data using fragile watermarks [A]. Proceedings of the 4th ACM Workshop on Digital Rights Management [C]. US: ACM, 2004. 73 – 82.

[41] BHATTACHARYA S, CORTESI A. A distortion free watermark framework for relational databases [A]. Proceedings of the 4th International Conference on Software and Data Technologies (ICSOFT'09) [C]. US: IEEE, 2009. 229 – 234.

[42] BHATTACHARYA S, CORTESI A. A generic distortion free watermarking technique for relational databases [A]. International Conference on Information Systems Security [C]. Berlin, Heidelberg: Springer, 2009. 252 – 264.

[43] LI Y. Database Watermarking: A Systematic View [M]. Berlin, Heidelberg: Springer Verlag, 2007.

[44] TSAI M H, HSU F Y, CHANG J D, et al. Fragile database watermarking for malicious tamper detection using support vector regression [A]. Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP'07) [C]. US: IEEE, 2007. 493 – 496.

[45] LI Y, DENG R H. Publicly verifiable ownership protection for relational databases [A]. Proceedings of the 2006 ACM Symposium on Information, Computer and Communications Security [C]. US: ACM, 2006. 78 – 89.

[46] HALDER R, CORTESI A. A persistent public watermarking of relational databases [A]. Proceedings of the 6th International Conferenceon Information Systems Security (ICISS'10) [C]. Gandhinagar, Gujarat, India: Springer LNCS, Volume 6503, 2010. 216 – 230.

[47] HALDER R, CORTESI A. Persistent watermarking of relational databases [A]. Proceedings of the IEEE International Conferenceon Advances in Communication, Network, and Computing (CNC'10) [C]. Calicut, Kerala,

India:IEEE Computer Society,2010.46 -52.
[48] BHATTACHARYA S,CORTESI A. A generic distortion free watermarking technique for relational databases[A]. Proceedings of the 5th International Conference on Information Systems Security (ICISS'09)[C]. Kolkata,India:Springer LNCS,Volume 5905,2009.252 -264.
[49] BHATTACHARYA S,CORTESI A. Database authentication by distortion-free watermarking[A]. Proceedings of the 5th International Conference on Software and Data Technologies (ICSOFT'10)[C]. Athens,Greece:INSTICC Press,2010.219 -226.
[50] KAMEL I. A schema for protecting the integrity of databases[J]. Computers & Security,2009,(28):698 -709.
[51] KHAN A,HUSAIN S A. A fragile zero watermarking scheme to detect and characterize malicious modifications in database relations[J]. The Scientific World Journal,2013,doi:10.1155/2013/796726:1 -16.
[52] LE H D,XU X,WANG Q,et al. Zero-watermarking for face image protection in database[J]. Journal of Internet Technology,2016,17(1):129 -135.
[53] POURNAGHSHBAND V. A new watermarking approach for relational data[A]. Proceedings of the 46th Conference on Annual Southeast Regional[C]. US:ACM,2008.127 -131.
[54] PRASANNAKUMARI V. A robust tamperproof watermarking for data integrity in relational databases[J]. Research Journal of Information Technology,2009,1(3):115 -121.
[55] LI Y,SWARUP V,JAJODIA S. Constructing a virtual primary key for fingerprinting relational data[A]. Proceedings of the 3rd ACM Workshop on Digital Rights Management[C]. US:ACM,2003.133 -141.
[56] LI Y,SWARUP V,JAJODIA S. Fingerprinting relational databases:Schemes and specialties[J]. IEEE Transactions on Dependable and Secure Computing,2005,2(1):34 -45.
[57] HALDER R,CORTESI A. Persistent watermarking of relational databases[A]. Proceedings of the IEEE International Conference on Advances in Communication,Network,and Computing (CNC'10)[C]. US:IEEE,2010.4 -5.
[58] ALFAGI A S,MANAF A A,HAMIDA B A,et al. A systematic literature review on necessity,challenges,applications and attacks of watermarking relational database[J]. Journal of Telecommunication,Electronic and Computer Engineering (JTEC),2017,9(1 -3):101 -108.
[59] 崔新春,贺洁,秦小麟. 基于盲源分离的多重音频数据库水印算法[J]. 电子学报,2012,40(1):78 -83.
CUI Xin-chun,HE Jie,QIN Xiao-lin. A multiple audio watermarking for database based on blind source separation[J]. Acta Electronica Sinica,2012,40(1):78 -83. (in Chinese)
[60] 付剑晶,王珂,徐建军. 一种面向多波段数字遥感影像的版权保护方案[J]. 电子学报,2016,44(3):732 -739.
FU Jian-jing,WANG Ke,XU Jian-jun. A copyright protection scheme for multiband digital remote sensing imagery[J]. Acta Electronica Sinica,2016,44(3):732 -739. (in Chinese)
[61] 王国栋,刘粉林,刘媛,等. 一种能区分水印或内容篡改的脆弱水印算法[J]. 电子学报,2008,36(7):1349 -1354.
WANG Guo-dong,LIU Fen-lin,LIU Yuan,et al. An image authentication scheme with discrimination of tampers on watermark or image[J]. Acta Electronica Sinica,2008,36(7):1349 -1354. (in Chinese)

作者简介

王春东 男,1969 出生于天津. 天津理工大学计算机科学与工程学院教授,博导. 1991 年于天津师范大学计算机科学专业获得工学学士学位. 分别于 2002 年和 2007 年在南开大学获得计算机科学硕士学位和博士学位. 主要研究方向为网络信息安全,普适计算,移动计算和智能信息处理.
E-mail:michael3769@163.com

杨 蕾 女,1993 年出生于河南漯河. 天津理工大学电子电气工程学院信息与通信工程专业研究生. 2016 年获得南京邮电大学理学硕士学位. 主要研究方向为信息安全,数据库水印,数据溯源,数据脱敏.
E-mail:guashushang89757@163.com