

基于混沌的 DCT 域关系数据库水印算法^{*}

周 飞, 赵怀勋

(武警工程学院 通信工程系, 西安 710086)

摘 要: 借鉴多媒体数字水印技术,提出了一种基于混沌的 DCT 域关系数据库水印算法,运用混沌序列对版权图像加密生成水印信息,并嵌入到候选属性的 DCT 中频系数中。利用 DCT 在频域中实现水印与载体的融合,隐蔽性更强,弥补了空域水印算法的不足;混沌序列的伪随机性和初值敏感性为数字水印的安全性提供了保证。实验表明,该算法具有较强的鲁棒性和良好的不可见性。

关键词: 关系数据库水印;离散余弦变换;混沌映射;鲁棒性

中图分类号: TP309.2 **文献标志码:** A **文章编号:** 1001-3695(2012)02-0786-03

doi:10.3969/j.issn.1001-3695.2012.02.104

Relational database watermarking algorithm based on chaos and DCT

ZHOU Fei, ZHAO Huai-xun

(Dept. of Communication Engineering, College of APF, Xi'an 710086, China)

Abstract: This paper presented a relational database robust watermarking algorithm based on chaos and DCT, which took the advantage of the frequency domain approach of multimedia digital watermarking into the area of relational database watermarking. By DCT on grouped and sorted tuples and modifying some of the transformed frequency coefficients to embed image watermark, only affecting one attribute for each tuple, the algorithm had good watermarking invisibility. Pseudo-random, sensitive to initial value of chaotic sequences improved the security of the watermark. The extraction did not need the original data. The experimental results show that the algorithm possesses strong robustness and good invisibility.

Key words: relational database watermarks; DCT; chaotic encryption; robustness

0 引言

目前,数据库水印技术有两个难题亟待解决:a)数据冗余小,数据库不同字段的属性要求不同,难以找到大规模的冗余空间;b)水印算法对鲁棒性要求很高,如果鲁棒性不强,不仅难以抵抗外部的恶意攻击,而且数据库的正常更新也会对嵌入的水印造成很大影响。在确保数据库数据的可用性和水印的不可见性的前提下,针对如何拓展冗余空间和如何提高算法鲁棒性两个问题,许多学者进行了大量的研究探索。Agrawal 等人^[1]提出的算法通过修改一些元组的数值型属性值的最低有效位(LSB)来实现水印信息的嵌入,该算法易于实现,但鲁棒性较差,抗攻击能力较弱,而且难以嵌入有实际意义的水印信息。2003 年 Sion 等人^[2]通过修改连续序列的数据分布特征来实现水印的嵌入,提出了一种鲁棒性数据库水印方案,但该方案会改变数据库的统计特征,增加数据更新开销。同年,牛夏牧等人^[3]进一步研究数据库水印,实现将有意义的水印信息嵌入数据库,但针对不同类型的数据库会产生病态结果。陈明刚等人^[4]构造主键的最小非负剩余系对数据分组,通过设置各组数据小波变换后“高/低频数据对”奇偶性的异同以嵌入信息,增强了水印的不可见性,但抵抗截尾攻击的能力较差。

针对上述问题,本文提出一种基于混沌的 DCT 域关系数据库水印算法,将版权图像进行混沌加密得到水印,利用主键哈希值选出有效的元组作为载体,对这些元组的候选属性进行

DCT,通过修改中频系数来嵌入水印信息。本文将水印算法与混沌理论相结合,在 DCT 域中嵌入水印,隐蔽性更强,弥补了空域水印算法的不足,加强了算法的鲁棒性。

1 基于混沌序列的加密算法

混沌序列既非周期性又不收敛,是一种具有时间遍历性、连续宽带频谱的随机过程,这种天然的隐蔽特性和对初始条件、参量动态的高度敏感性,使得混沌序列非常适合用于加密水印信息。Logistic 映射是典型的、也是应用最广泛的混沌模型。由 Logistic 映射生成的混沌序列主要具有以下优点^[5,6]:

a)改变混沌系统的参数及初值,即可得到数量极多、长度任意的序列;

b)序列的概率密度分布函数与初值无关,类似于一个随机过程,保密性良好;

c)序列的产生复制非常方便,只需给出一个混沌迭代式和初值,即可产生一个混沌序列,不必花费空间存储整个序列。

Logistic 映射定义为: $X_{n+1} = \mu X_n (1 - X_n)$ 。其中, $X_n \in (0, 1)$,当 $3.5699456 < \mu \leq 4$ 时 Logistic 映射呈现混沌状态,即由初始状态 X_0 生成的序列 $\{X_j, j=0, 1, 2, 3, \dots\}$ 是既非周期又不收敛且不相关的,并对初值非常敏感。

本文采用的方法是将作为版权标记的二值图像进行混沌二值序列调制,再用混沌序列加密后作为水印信息嵌入数据库中,这样有效地加强了版权信息的保密性。Logistic 序列的遍

收稿日期:2011-06-13;修回日期:2011-07-26 基金项目:陕西省自然科学基金研究计划基金资助项目(2007F50)

作者简介:周飞(1987-),女,河北宁晋人,硕士研究生,主要研究方向为数字水印、信息安全(zhoufei_wj@163.com);赵怀勋(1960-),男,教授,主要研究方向为网络与信息安全技术。

历统计特性等同于零均值白噪声,随机性、相关性和复杂性能良好。因此用 Logistic 混沌序列对版权图像进行预处理,不仅能增强水印信息的保密性,而且能有效提高抵抗图像剪切攻击。

2 基于混沌的 DCT 域关系数据库水印算法

2.1 研究前提

假设在数据使用范围内,对某些数据进行的微小改动不会影响数据的使用。关系数据表中的某数值型属性能够容忍由于嵌入水印而带来的微小变化,则称该属性为候选属性。候选属性的取值为实数,对其中的一些数位进行微小改动不会影响数据效用的,称为最不重要位(least significant bit, LSB)。另一些因为微小改动而极大影响数据效用的数位,称为最重要位(most significant bit, MSB)。LSB 一般取数值的末尾几位,而 MSB 一般取首位。另外,假设数据库中元组的主键是不能被修改的,文献[7]就是着重处理这一问题的。

在不失一般性的情况下,设数据库关系为 $R(P, A_1, A_2, \dots, A_n)$, 其中 P 为主码, $A_j(1 \leq j \leq s)$ 是第 j 个数值型属性。水印算法中用到的参数和符号如表 1 所示。

表 1 水印算法中的符号及其含义

符号	含义
n	R 中元组数
s	R 中数值属性的个数
g	R 中分组数
g_i	第 i 个分组
m	一个分组的元组个数
r_i	R 中或分组中的第 i 个元组
$r_i.A_j$	R 中或分组中第 i 个元组的第 j 个属性
$r_i.P$	R 中或分组中第 i 个元组的主键
K_1, K_2, K_3	密钥
W	嵌入数据库中的水印
W'	检测出的水印
w_i	水印的第 i 位
$r_i.hash_key$	第 i 个元组的哈希值
$r_i.hash_mod$	第 i 个元组的分组号
$r_i.location$	第 i 个元组嵌入水印的标记属性位置
$r_i.coefficient$	可嵌入水印的第 i 个元组的系数
k_j	第 j 个属性允许修改的最大位数
c_j	DCT 的第 j 个中频系数

2.2 嵌入算法

为了同时实现不可见性和鲁棒性,将版权图像预处理后嵌入到数据 DCT 后的中频部分,如图 1 所示。首先将要嵌入版权信息图片转换为二进制比特流,设比特流为 $I(i)(i=1, 2, \dots, u \times v)$ 。同时,混沌序列的散列处理给定混沌初值(密钥 K_1)生成混沌随机序列 $L(i)$,截取混沌随机序列 $L(i)$ 长度 $u \times v$,生成加密处理后的待嵌入水印信息 $W(i) = I(i) \oplus L(i)(i=1, 2, \dots, u \times v; \oplus$ 为异或运算)。

根据属性值 A 、元组主码 P 和用户密钥 K ,计算元组哈希值,对元组进行分组,并对每组进行排序。根据属性允许修改的最大位数 k_j 和元组哈希值,选择水印嵌入属性位置,并确定将要进行 DCT 变换的属性值 $r_i.coefficient$,得到一组数据集。

采用零填充和截断方法对数据集预处理,使得表示方法标准化,再进行正向离散余弦变换(FDCT)。由于变换得到的高频系数容易遭到低通滤波的攻击,而低频系数集中了大部分能量,因此为了保持不可见性和鲁棒性之间的平衡,选择在中频系数中嵌入水印信息。具体实现过程如图 1 所示。

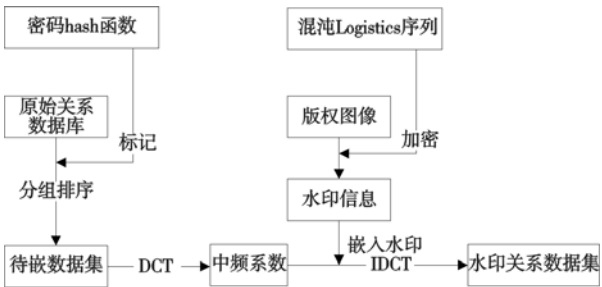


图1 水印嵌入过程

每个分组中嵌入 1 bit 水印,具体的嵌入策略可以根据数据库的数据特点选择适当的方法。本文在仿真实验中采用的方法是对中频系数 c_j 进行如下修改:if $(w_i == 1 \& c_j > 0 \parallel w_i == 0 \& c_j < 0) c_j = -c_j$ 。算法可以根据对鲁棒性的要求来调节修改 c_j 的数量。修改的 c_j 越多,算法鲁棒性越强;反之,对数据库影响越小,水印的不可见性越好。

算法 1 水印嵌入算法

```
for i=0 to M-1 do//M 为版权图像转换成二进制流长度
    encode(1); //利用混沌序列对版权信息加密
for each tuple  $r_i$  in the relation
     $r_i.hash\_key = hash(K_2, r_i.P, r_i.A_j)$ ; //计算元组哈希值
     $r_i.hash\_mod = mod(r_i.hash\_key, g)$ ; //分组,  $g = ceiling(n/m)$ 
for each group  $g_i$  in the relation
    sort( $g_i$ ); //对每个分组进行排序
for each group  $g_i$  in the relation
    for each tuple  $r_j$  in  $g_i$ 
         $r_j.location = location(r_i.hash\_key, k_p) \% s$ ;
        //选择嵌入水印的属性位置  $(1 \leq p \leq s)$ 
         $r_i.coefficient = select(r_i.P, r_i.location, k_p)$ ;
        //确定参与 DCT 的属性值
    pretreatment(coefficient_list); //对数据集进行预处理
    DCT.Fdct(coefficient_list); //FDCT 正向离散余弦变换
    Watermark_embed( $W[q + \% W.length]$ , coefficient_list, times);
    //在分组中嵌入 1 bit 水印
    DCT.Idct(coefficient_list); //逆 DCT 处理
    Posttreatment(coefficient_list); //压缩数据量
    Update(database); //更新数据库
```

2.3 检测算法

水印检测过程与嵌入相对应,如图 2 所示,利用多数选举法提取水印信息,再将水印信息还原成版权图像。在仿真实验中采用的具体策略是:

```
if (count( $c_j > 0$ ) > count( $c_j < 0$ ))  $w[q] = 0$ ;
else if (count( $c_j > 0$ ) < count( $c_j < 0$ ))  $w[q] = 1$ 。
```

为了判定是否有水印存在,需要将所提取出的版权图像 M^* 与原始版权图像 M 进行如下相关检测:当 τ 大于阈值 τ_0 时,水印信息存在;否则,水印信息不存在。阈值 τ_0 可根据实际情况设定,需要考虑数据库的正常更新操作以及可能遭到的恶意攻击,即算法的鲁棒性和准确性的均衡。

3 实验仿真及算法分析

为检验算法的有效性及相关性能,利用某城市街道规划数据库的数据进行了相关实验。该关系数据库中共有 100 000 个元组,每个元组有 31 个属性,选取其中的 4 个数值型属性嵌入水印,实验环境为 MATLAB 7.0,版权图像采用 64×64 的二

值图像。为了不失一般性,假设属性值允许修改最多位数为3位,即 $k_p=3(1\leq p\leq s)$ 。实验中,每组修改 c_j 的个数为15个。另外,实验中DCT采用 8×8 浮点运算,若将要进行DCT的元组属性值不足64位则补0,多于64位则取前64位进行变换运算。

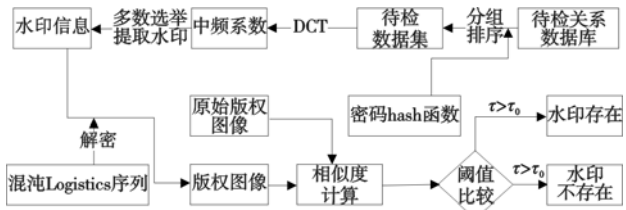


图2 水印检测过程

3.1 算法分析

1) 复杂度分析

算法对数据的操作包括数据读写操作、分组、排序、DCT和逆变换以及水印嵌入和检测。算法的复杂度主要取决于分组、排序和DCT,其时间复杂度分别为 $O(n)$ 、 $O\left(n\log\left(\frac{n}{g}\right)\right)$ 和 $O\left(n\log\left(\frac{n}{g}\right)\right)$,因此算法的时间复杂度为 $O\left(n\log\left(\frac{n}{g}\right)\right)$ 。

2) 鲁棒性分析

算法引入混沌系统对版权信息进行预处理,去除了图像的自相关性,各像素点接近于随机分布,不仅提高了水印的抗攻击能力,即使数据库在使用过程中水印遭到破坏,经过解密变换也能够把损坏的比特分散开来,降低对视觉的影响;而且增强了版权信息的隐蔽性,即使水印被截获,如果不能正确解密,仍然无法得到水印信息。此外,水印的生成、嵌入与检测过程均在密钥控制下进行,有效提高了算法的安全性。

3) 不可见性分析

算法利用主码的hash值对数据进行分组、排序,使得水印信息在各分组中均匀分布,提高了水印的不可见性;对分组内部分数据进行DCT,将水印嵌入中频系数,对数据的整体分布特性影响较小,使得水印信息有较好的隐蔽性。

3.2 实验结果

1) 水印嵌入误差

从水印嵌入前后的属性统计特征来看,本文算法将数字水印嵌入到关系数据DCT的中频系数,使得水印信息更加均匀、分散,得到了良好的隐蔽性。水印信息嵌入前后的统计信息如表2所示,可以看出由水印引起的误差非常小。

表2 水印对数据库的整体影响

属性名称	均值	方差	均值变化 (10^{-9})	方差变化 (10^{-6})
属性1	174.7838747873	208.4995268776	1.5	6.3008
属性2	301.8750275186	23.8594676104	0.37	3.0416
属性3	114.0572467825	214.2625229440	0.83	4.3505
属性4	134.1217749758	148.5049761206	0.26	2.1288

2) 抗攻击性能

实验模拟攻击者对算法进行子集选择、添加、修改攻击,得到的仿真结果如图3所示。由图3可知,对于子集选择攻击,选择的数据比例越大,相应提取水印信息就越多,准确恢复水印的概率就越高;对于子集添加和修改得越多,水印的提取率就越低。

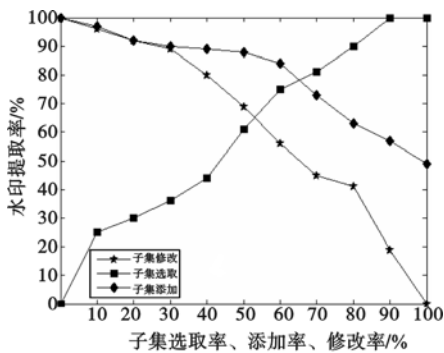


图3 子集选取、添加、修改攻击

攻击者一般不会对重要数据进行较大幅度的修改或删除,否则该数据库就失去了其使用价值。本文对数据库进行了多次随机更改模拟攻击,实验结果如表3所示。从实验结果可以看出,当数据修改达到40%时,版权信息还是可以辨认出来。

表3 子集修改攻击检测实验结果

版权图像	检测出的版权图像					
	500	1000	5000	10000	20000	40000
工程学院	工程学院	工程学院	工程学院	工程学院	工程学院	工程学院

4 结束语

本文借鉴多媒体数字水印技术,设计了一种基于混沌的DCT域关系数据库水印算法。该算法利用混沌序列对版权图像低频加密生成水印,并嵌入到部分候选属性的DCT中频系数中,避免了对数据库的大幅修改,对关系数据库的均值和方差的影响都很小。理论分析和实验结果证明该算法不可见性好,具有较强的抗攻击性。

参考文献:

[1] AGRAWAL R, KIERNAN J. Watermarking relational databases [C]// Proc of the 28th VLDB Conference. 2002:155-168.

[2] SION R, ATALLAH M, PRABHAKAR S. Rights protection for relational data [J]. IEEE Trans on Knowledge and Data Engineering, 2004, 16(12): 1509-1525.

[3] 牛夏牧,赵亮,黄文军. 利用数字水印技术实现数据库的版权保护 [J]. 电子学报, 2003, 31(B12): 2050-2053.

[4] 陈明刚,孙星明,肖湘蓉. 基于小波变换的关系数据库水印算法 [J]. 东南大学学报:自然科学版, 2007, 37(1s): 130-136.

[5] 王东生,曹磊. 混沌、分形及其应用 [M]. 合肥:中国科学技术大学出版社, 1995: 1-192.

[6] FENG Guo-rui, JIANG Leng-ge, HE Chen, et al. Chaotic spread spectrum watermark of optimal space-filling curves [J]. Chaos Solitons and Fractals, 2006, 27(3): 580-587.

[7] 向玥,李军义,潘季芳. 一种基于虚拟主键的数据库水印算法 [J]. 计算机研究与发展, 2009, 46(Z1): 66-70.

[8] CAO Zai-hui, SUN Jian-hua, HU Zhong-yan. Image algorithm for watermarking relational databases based on chaos [C]//Advances in Wireless Networks and Information Systems. Heidelberg: Springer, 2010: 411-418.

[9] FARFOURA M E, HORNG S. A novel blind reversible method for watermarking relational databases [C]//Proc of International Symposium on Parallel and Distributed Processing with Applications. 2010: 563-569.