

一种关系数据库的多媒体多重水印

赵学民, 李玲玲

(郑州航空工业管理学院 计算机科学与技术系, 河南 郑州 450015)

摘 要: 针对关系数据库水印算法的一些不足, 提出了一种保护关系数据库的多重水印方法. 该方法将密钥和语音等多媒体信号结合在一起作为水印嵌入关系数据库, 实现了多重水印在关系数据库中的嵌入和提取, 并且多重水印提取时不需要原始数据的参与. 实验表明, 该方法有效地平衡透明性与稳健性之间的关系, 对于子集选择攻击和子集改变攻击具有较好的稳健性.

关键词: 数据库; 数字水印; 数据库水印; 信息隐藏

中图分类号: TP311

文献标识码: A

文章编号: 1000-7180(2013)09-0122-04

A Multiple Watermarking of Multi-Media for Relational Databases

ZHAO Xue-min, LI Ling-ling

(Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou 450015, China)

Abstract: A multiple watermarking algorithm is proposed to protect relational databases copyright for some lacks of the existing multi-watermarking algorithms. The novel multiple watermarking scheme, which embeds multi-media watermarks into relational database. The watermark embedding and extraction algorithms are specified. The algorithm is "blind" in that it isn't requires original data in order to detect a watermark. The results of the corresponding watermark experiments and the attack experiments verify that the proposed method is correct feasible and robust.

Key words: database; digital watermarking; database watermarking; information hiding

1 引言

随着数据库技术的不断发展及数据库管理系统的广泛应用, 数据库的安全问题越来越受到重视. 目前数据库安全方法主要可以通过访问控制、数据加密等技术实现, 如银行、证券、通信、医疗等数据. 然而, 利用数字水印技术采用非加密的方法, 实现对数据库的版权保护和安全控制, 更符合关系数据的特性和应用需求. 与多媒体数据相比, 数据库具有规范的结构和语义、数据量大、经常需要更新, 且对数据的微小改动都有可能影响数据的可用性, 可用于嵌入水印的冗余信息较少. 由于关系数据库本身的特殊性, 在保证数据可用性的前提下, 将版权信息嵌入到数据库中难度较大, 传统的多媒体数字水印技术

难以直接应用到其中.

目前, 人们在积极探索如何使信息隐藏起来使其避免受到有意或无意攻击, 从而寻求对攻击具有免疫力的数据库水印嵌入算法, 主要研究如下:

(1) 数据库水印的鲁棒性及安全性研究, 主要应用于数字产品的版权证明. Francesc Sebe 等人^[1]提出了如何保持经过加水印后的数据库数据的平均值, 并且对于子集添加攻击有着很好的鲁棒性的水印算法. 牛夏牧等人^[2]通过在嵌入水印前先对数值的变化范围设置一个约束限制条件, 然后进行筛选, 利用满足约束条件的属性值的最低有效位来嵌入水印信息, 实现了在多个属性集中嵌入多位水印信息. 从而来对数据库的版权进行验证. 张勇等人^[3]结合云模型理论来进行关系数据库水印的嵌入以及提取.

收稿日期: 2012-11-08; 修回日期: 2012-12-14

基金项目: 国家自然科学基金(41171341); 河南省科技厅科技攻关项目(102102210453)

(2)数据库脆弱水印方面的研究,主要用来实现对数据的完整性验证. Yingjiu Li^[4]等人最早对关系数据库脆弱性水印算法进行了研究,实现了数据库的完整性验证. 肖湘蓉^[5]等人利用数值数据在一定范围内是允许变化的特征来构造脆弱性水印,能够很好地判定水印攻击和对数据库进行恢复.

(3)水印宿主数据类型扩展方面的研究. Radu Sion 等人^[6]首先提出了利用字符型数据的分布特性和属性与主键值之间关系的冗余空间来实现非数值型数据的水印信息嵌入,并用实验证明了该方案有良好的鲁棒性.

(4)数据库数字指纹研究. Camelia Constantin 等人^[7]研究了在保持数据完整性约束的条件下适用于数据库和 XML 文档的通用指纹算法. Yingjiu Li 等人^[8]提出了一种新的数据库数字指纹嵌入模型,并通过实验对这一模型的鲁棒性进行了分析.

本文提出了一种使用多媒体信息对关系数据库进行多重水印的方法. 该方法首先将一个标志图像嵌入关系数据库代表版权信息,同时将语音作为最初的水印嵌入. 因为对人类来说,语音是不会随着时间改变的. 所以用语音作为水印可以弥补上述的限制. 实验证明,这种脆弱水印是有效的和具有说服力的,水印嵌入过程是不易察觉和安全可靠的,该方法有效地平衡透明性与稳健性之间的关系,对于集选择攻击和子集改变攻击具有较好的稳健性.

2 水印系统的结构

在该方案中,用版权持有人的语音和其他信息通过水印生成算法生成水印,并给出相应的插入算法和检测算法(见图1). 关系数据库水印问题可以被精确地模拟如下:假设关系 R 包含主键 P 和数值属性 $A_0, A_1, \dots, A_{v-1}$. 假定它是接受变化的 ξ 最小有效位(LSB)之一. 一个字符图像可以转换为0和1的序列,这些序列被嵌入了以版权保护为目的的关系 R ^[3].

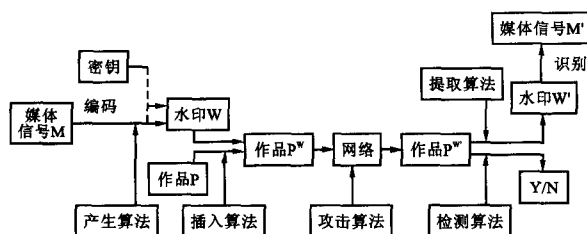


图1 水印系统框架

在该水印系统框架结构中,第一步是水印的产

生,语音等媒体信号 M 首先进行编码,编码和密钥经过产生算法的运算生成水印 W ;第二步是水印的嵌入,由插入算法将水印 W 嵌入到要保护的作品 P 中,得到作品 P' ,作品 P' 通过网络分发给合法的被授权者,如果受到非法复制和篡改,即是经过攻击算法的运算生成作品 P'' ;第三步水印的提取和检测,如果作品的所有者怀疑其受一些非法复制或损坏,可以用检测算法检测或者用提取算法提取出 P'' 中的水印 W' 和媒体 M' ,从而知道版权作品有没有被非法拷贝或篡改.

3 水印算法的实现

关系数据库中属性值的一些微小变化是可以接受的,从而将版权信息嵌入到这些属性中. 可以把代表版权信息的字符图像看成是0和1的序列,0和1的标记是关系数据中只是一些小的错误. 所有0和1的标志代表整个版权信息.

3.1 水印生成算法

水印生成算法包括媒体信号压缩和编码两个步骤. 一是媒体信号压缩. 水印关系数据库通过嵌入水印引入一些小的标记. 这些小的标记不会显著影响有用的数据,因此水印应当很小. 因此,压缩图像信号是必要的. 由于图像信息具有很大的信息冗余,可以使用小波变换进行压缩. 二是信号编码. 将语音信号波形转换为8位A律编码,版权图像像素是按照先从左到右,然后从上到下顺序安排的,从而可以得到一个集合 $S = \{ S_1, S_2, \dots, S_{N \times N} \}$,其每一个元素是一个十进制的像素灰度值.

3.2 水印嵌入算法

在水印嵌入时,使用一个单向散列函数来决定主键 P 和 K 密钥应该在哪里标记和标记什么. 为了在关系型数据库中嵌入水印需要做两个基本假设:一是水印嵌入的对象关系数据库只有数值属性,二是关系数据库的数值属性值能够容忍小范围的变化. 基于上述两个基本假设,能够在对数据最小影响的情况下实现嵌入. 嵌入程序如下:

```
// Weighted watermark algorithm, return marked R
// Hi is one-way hash functions, L is the length of EMC
// Parameters k, L,  $\alpha$ ,  $\xi$  and  $\gamma$  are private to the owner.
1 E[L] = H(k concatenate M) // calculate L-bit EMC
2 for each tuple r  $\in$  R do
3   t = H1(k concatenate r, P)
4   if ( t mod  $\gamma$  equals 0) then // mark this tuple
5     i = select_attribute() // mark i-th attribute
6     j = t mod  $\xi$  // mark j-th bit
```

```

7      k= t mod L      //use the k-th bit of EMC
8      m=Ek XOR (k mod 2) // value of marked bit
9      set the j-th LSB of r. Ai to m
10     return R
11     procedure select_attribute()
12         u= H2( k concatenate r. P)
13         d= Σ(1/ Wi) i∈[0,i-1]
14         if (u mod d) ∈ (Σ(1/ Wi-1), Σ(1/ Wi])
15             then return i // W-1=0

```

3.3 水印提取算法

如果数据的所有者怀疑其关系数据库中的一些数据受到了非法复制或损坏. 他或第三方可以使用水印提取算法来验证可疑数据的所有权. 下面介绍该水印提取算法.

因为相同的键作为哈希函数的种子会产生相同的输出, 每一组中被选者的元组在插入算法中具有相同的秩序. 对于每一个标志位, 我们可以根据其价值的数字分别地将其作为 0 或者 1, 然后根据多数表决机制来决定这一位的最后值. 检测结果是一个包括版权信息的数据集二进制序列. 因此, 我们可以把二进制序列转换成版权的水印图像. 著作权人可以利用水印图像来证明它的版权. 水印检测算法如下:

```

// Algorithm to return a watermark M[ ] from relation R
// parameters k, L, α, ξ and ν are also private to the owner.
1  for s=0 to L-1 do
2      DM[s]='' // initialize detected mark code
3      count[s][0]=0, count[s][1]=0 // initialize counter
4      for each tuple r∈R do
5          t= H1( k concatenate r. P)
6          if ( t mod γ equals 0) then // select this tuple
7              i = select_attribute() // mark i-th attribute
8              j = t mod ξ // select j-th bit
9              k= t mod L // mark the k-th bit of EMC
10             m= ( j-th LSB of r. Ai) XOR (k mod 2)
11             count[k][m]=count[k][m]+1 // add the counter
12             for s=0 to L-1 // get the watermark
13                 f (count[s][0]>=count[s][1])// majority voting
14                 then M[s]=0 else M[s]=1 //the final bit value

```

4 实验及其结果

为了测试算法的有效性和鲁棒性, 在 2.4 GHz CPU 和 256MB 内存运行 Windows XP 专业版的计算机上进行了实验. 实验平台是基于 JDBC 的 Java Eclipse 可视化平台和 Microsoft Office Access 2003. 测试时, 算法生成了具有 8 种属性的合成数

据, 数据集的大小是 30000, 选择了 MD5 作为单向散列函数, 显著水平是 0.01. 基于图像的关系数据库多重水印图像, 每一个水印图像都有它自己的参数, 用于插入算法和相应的提取算法. 在实验中, 测试了在关系型数据库中多种攻击情况, 下面仅给出子集选择攻击和子集变更攻击的测试结果和分析.

(1) 子集选择

攻击者仅选择关系 R 的一个子集. 在实验中, 选择了不同比率的原始数据来模拟这种攻击^[6]. 在实验中, 选择水印关系的 50% 对整个水印基本没有影响, 即使选中 25% 的数据, 水印图像的轮廓制图也可以成功恢复. 实验结果显示, 即使很小一部分关系被标记也可以成功的检测出来. 实验结果如图 2 所示.

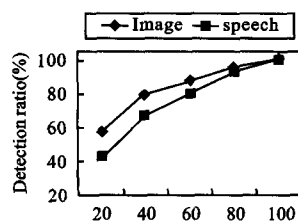


图2 子集选择攻击的结果

(2) 子集变更

所谓子集变更攻击即是攻击者试图改变关系数据库水印的属性值或推测性地变更某些关系的属性值. 实验中, 通过在每一元组的 8 位中重置相反的 1 位, 即把值为 0 变为 1 把值为 1 变为 0, 来随机改变元组的一部分. 结果显示, 当改变 50% 或更少的元组时, 可以成功地恢复整个水印图像. 当 60% 或更多的元组受到攻击时, 就不能检测到真个水印图像. 实验结果, 如图 3 所示.

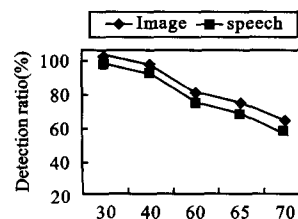


图3 子集改变攻击的结果

该算法是一个盲算法, 它不需要原始数据和水印来检测一个目标的水印. 这样, 不管是在关系数据库的一份拷贝还是原来数据库的更新中, 都可以检测到水印, 并且水印还可以简单有效地保持.

5 结束语

本文提出了一个利用媒体作为水印的关系型数

据库水印的新方法. 通过实验表明,当攻击者漏掉一些元组或者关系中的水印属性时,该算法仍然有良好的鲁棒性. 而且,根据省略属性的数量和漏掉元组的数目,能够估计出检测水印所需的样本数. 该方法较为直观,水印容易鉴定.

参考文献:

- [1] Francesc Sebe, Josep Domingo-Ferrer. Agusti solanas noise-robust watermarking for numerical datasets[C] // Proceeding of the MDAI 2005, Berlio Heidelberg: Springer-verlag, 2005: 134-143.
- [2] 牛夏牧, 赵亮, 黄文军, 等. 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(12): 2050-2053.
- [3] 张勇, 赵东宁, 李德毅. 关系数据库数字水印技术[J]. 计算机工程与应用, 2003(25): 193-195.
- [4] Yingjiu Li, Huiping Guo, Sushil Jajodia. Tamper Detection and Localization for Categorical Data Using Fragile Watermarks[C] // Proceeding of the DRM'04, Washington, DC, USA: IEEE, 2004: 210-221.
- [5] 肖 湘蓉, 孙星明. 基于水印的数据库安全控制研究[J]. 计算机工程与应用, 2005, 41(6): 175-178.
- [6] Radu Sion, Mikhail Atallah, Sunil Prabhakar. Ownership proofs for categorical data [J]. IEEE Journal of Transactions on Knowledge and Data Engineering IEEE TKDE, 2005, 17(7): 223-225.
- [7] C Constantin, D Gross— Amblard et M Guerrouani. Watermill: an Optimized Fingerprinting Tool for Highly Constrained Data [C] // ACM Workshop on Multimedia and Security(MMSEC), New York, USA: IEEE, 2005, (08): 143-155.
- [8] Yingjiu Li, Vipin Swarup. Fingerprinting relational databases schemes and specialties [J]. IEEE Transaction on Dependable and Secure Computing, 2005, 2(1): 34-45.

作者简介:

赵学民 男, (1977-), 硕士, 讲师. 研究方向为网络信息安全等.

李玲玲 女, (1973-), 博士, 副教授. 研究方向图像处理等.

(上接第 121 页)

一种基于嵌入式 RAM 单元的实现结构, 叙述了设计到优化实现的过程. 该结构采用 CIC 抽取滤波器与 CIC 补偿滤波器级联的方式, 实现时利用 RAM 单元来完成数据缓存, 在达到节省 DSP 单元的目的的同时, 大量减少了逻辑单元消耗. 本文设计的滤波器实现方法具有较强的通用性.

参考文献:

- [1] 杨小牛, 楼才义. 软件无线电原理与应用[M]. 北京: 电子工业出版社, 2001.
- [2] 何建标, 王宏远, 郭跃, 等. 一种基于 FPGA 的 FIR 滤波器实现结构[J]. 微电子学与计算机, 2008, 25(3): 47-50.
- [3] Meyer-Baese U. 数字信号处理的 FPGA 实现 [M]. 刘 凌, 译. 北京: 清华大学出版社, 2011.
- [4] Hogenauer E B. An economical class of digital filters for decimation and interpolation [J]. IEEE Trans on Acoustics, Speech and Signal Processing, 1981, 29(2): 155-162.
- [5] 万能, 王江宏, 郭亮, 等. LTE 中变速率 FIR 滤波器的 FPGA 实现[J]. 微电子学与计算机, 2009, 26(11): 63-69.
- [6] Altera Corp. Cyclone III device handbook [M]. San Jose: Altera Corporation, 2008.

作者简介:

王亚磊 男, (1988-), 硕士研究生. 研究方向为数字信号处理、无线通信.