一种用于关系数据库的可逆水印技术

张 勇[1,2],牛夏牧[1,2]
(1. 哈尔滨工业大学深圳研究生院,广东深圳 518055;2. 深圳国际技术创新研究院,广东深圳 518055)

摘 要: 针对关系数据库数据在使用时必须要求真实性的原则,提出一种可用于关系数据库的可逆水印技术方案,该方案可以从无篡改的水印数据库中无损地恢复原始数据,该水印算法只有拥有密钥和其他秘密参数的用户才可以完全恢复原始数据.通过算法分析和实验分析,该水印方案还具有不可见性和可行性等特点.

关键词: 可逆水印; 数字水印; 关系数据库
中图分类号: TP391.41 **文献标识码**: A **文章编号**: 0372-2112 (2006) 12A-2425-04

Reversible Watermark Technique for Relational Databases

ZHANG Yong[1,2],NIU Xia-mu[1,2]
(1. *Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, Guangdong* 518055, *China*;
2. *Shenzhen Innovation International, Shenzhen, Guangdong* 518055, *China*)

Abstract: Now people are concerned about that how to protect the digital media copyright more and more. And for relational data, that how to protect its ownership and not to affect its usage becomes a difficult problem. A reversible watermarking scheme for relational databases on numerical attribute is proposed. The basic idea of the scheme is to embed the watermark information into relational databases with exclusive or under some secret parameter, and the original data can be recovered losslessly from the unmodified watermarked relational databases with exclusive or under the same secret parameter. Only the person who owns the private key and other unique parameters can recover them perfectly. Through the algorithms and some experiments analysis, the scheme is proved to be invisible and feasible.

Key words: reversible watermark; digital watermark; relational databases

1 引言

随着信息技术和网络技术的飞速发展,传统的模拟媒介正逐渐被数字媒体所代替[1],因此人们之间相互交流、传输和拷贝数字材料变得越来越容易.与此同时,人们也越来越关心他们数字财产的所有权问题.为了解决这个问题,20 世纪 90 年代初期出现了数字水印技术.根据不同的应用背景,数字水印可以分为鲁棒水印和脆弱水印等.由于嵌入载体的水印无容置疑地会或多或少地破坏所保护的对象本身,所以找到一种可从未篡改水印数据中完全分离出水印信息并可以恢复原始数据的可逆水印方案势在必行.为了适应关系数据库的安全需要,结合数据库数据的特点,本文提出了一种用于关系数据库的可逆水印技术方案.

2 相关工作

文献[2]提出并介绍了可逆水印技术的概念,紧接着,文献[3～9]提出了多种可逆水印技术方案,大多应用于多媒体数据,而用于关系数据的可逆水印技术方案还未有公开文献出版.

文献[10～15]提出了几种用于解决关系数据库版权保护问题的水印技术方案,这些方案都是建立在关系数据有足够的冗余空间并且可以容忍一定的误差的基础上,所以嵌入的水印信息将修改数据的数值,其目的是通过检测和验证关系数据的版权从而保护所有者的所有权.文献[16]提出了一种脆弱水印技术方案,用于检测和定位关系数据是否篡改和篡改的位置,该方案中,尽管嵌入水印后的数据失真不容易被察觉,但毫无疑问载体数据也有失真的存在.以上提到的水印技术方案,都假定一些属性值的微小改变是可以容忍的.但是对于一些数据的应用不允许有任何的变化,而这些数据同样需要版权保护.为了解决这个问题,论文提出了一种适用于数值型属性数据的可逆水印技术方案.

3 水印算法

需要说明的是,为了增强所提出可逆水印技术方案的可

收稿日期:2006-04-10;修回日期:2006-11-24
基金项目:国家自然科学基金(No. 60372052,No. 60671064);广东省自然科学基金(No. 05109511);全国百名博专项资金 (No. FANEDD-200238);哈尔滨工业大学交叉性学科基金(No. HIT. MD-2002. 11);哈尔滨工业大学校基金(No. HIT. 2003. 52);黑龙江省优秀杰出青年基金;新世纪优秀人才计划(No. NCET-04-0330);国家 863 计划(No. 2005AA733120)

行性和不可见性,载体数据应有足够的冗余位置允许嵌入标记信息. 本文涉及到的数据库数据都假定可嵌入水印标志的冗余空间信息量远大于所要嵌入标记信息的信息量.

3.1 相关符号及其含义

水印算法中将用到的一些符号及其含义如表 1 所示. 这里,1/μ表示嵌入比例,意思是说如果载体数据有 x 个嵌入位置,那么就有大约 x/μ 个嵌入位置有待嵌入.

表 1 符号及其含义

符号	含义	符号	含义
R	关系数据库	A_i	R 中的第 i 个数值型属性
W	水印信息,也是所有者拥有的私钥	r	R 中的元组
m	水印信息的长度	$r.A_i$	元组 r 的第 i 个数值型属性的值
b	水印信息的二进制序列	c	$r.A_i$ 的二进制序列
b_i	b 的第 i 位信息	c_j	c 的第 j 位
n	b 的长度	Pk	R 的主键值
k	R 中用于水印操作的数值型属性的数目	μ	1/μ是嵌入比例,μ是一个正整数
j	$r.A_i$ 的嵌入位置		

3.2 可逆水印算法

可逆水印算法如图 1 所示. 在嵌入水印标记之前,应该确定载体数据的最大允许误差范围和待嵌入的位置 j. 对于 j 可以这样来确定,假如某个载体数据最大允许误差范围是 $(-10,10)$,可以确定嵌入位置 j 应为 3,这是因为 10 这个数大于 8(二进制序列为 1000,即 2^3),而小于 16(二进制序列为 10000,即 2^4).

```
1     将水印信息 W 转换为对应的二进制序列 b,假定 b 的长度是 8×m,记做 n,m 是水印信息 W 的长度;
2     确定最大允许误差范围,并确定二进制序列待嵌入的位置,记为 j;
3     For 每一个元组 r∈R
4     For 每一个属性 A_i∈R
4.1     If Hash(W°Pk°A_i) mod μ = 0 then
4.1.1     转换 r.A_i 为二进制序列 c;
4.1.2     s = Hash(Hash(W°Pk°A_i)) mod n;
4.1.3     if (c_{j+1} = '1') && ((b_s = '1') or (b_s = '0')) then
4.1.4       c_j = c_j XOR b_s;
4.1.5       将 c 十进制化并写入当前元组当前属性 r.A_i;
4.1.6     endif;
4.2     Else
4.2.1     If Hash(W°Pk°A_i) mod μ = 1 then
4.2.2       转换 r.A_i 为二进制序列 c;
4.2.3       s = Hash(Hash(W°Pk°A_i)) mod n;
4.2.4       if (c_{j+1} = '0') && ((b_s = '0') or (b_s = '1')) then
4.2.5         c_j = c_j XOR b_s;
4.2.6         将 c 十进制化并写入当前元组当前属性 r.A_i;
4.2.7       endif;
4.3       endif;
4.4     endif;
5     endfor;
6     endfor.
```

图 1 可逆水印算法

在图 1 中,符号'°'是一个连字符号,第 4.1 行表示通过连字符号'°'将当前元组中的 Pk 值,A_i 属性名,及 W 连接为一个字符串,然后通过单向 Hash 函数(如 MD5 和 SHA 等)转换为一个 128 位的二进制序列[13~15].

这里,W 也是所有者唯一拥有的密钥信息. 只有拥有密钥 W,嵌入比例 1/μ和嵌入位置 j 的所有者或者版权人才可以通过图 1 的算法很好地准确恢复出未篡改水印关系数据的原始数据

4 实验和算法分析

在上一节中,参数 W,Pk,μ and j 在可逆水印关系数据库算法中非常重要,并且算法还依赖于单向 Hash 函数(如 MD5 or SHA 等).

现在通过一些水印实验来检验 3.2 节中可逆水印算法的可行性和不可见性. 论文选择了一个随机生成的关系数据库作为载体数据,该数据库中有一个表,表有长、宽、高三个数值型属性,该表的元组数是 20,000 条. 这里,选定"mark"作为水印信息 W,嵌入比例分别为:0.1,0.3,0.5 和 1,$(-10,10)$是最大允许误差范围,从而可确定嵌入位置为 3(即 $j=3$). 相关实验结果如图 2 和图 3 所示,其中图 2 表示不同嵌入比例条件下所嵌入的"mark"对应的二进制序列中每一个二进制值嵌入的次数,从该图可以看出水印信息即使在 10 %的嵌入比例条件下,水印信息的二进制值都被嵌入到载体数据中的最小次数是 2,所以可以很方便地还原原始载体数据. 图 3 所示为原始数据,水印数据和还原后数据分布(这里的嵌入比例是 0.5). 从图 3 可以看出,通过可逆水印算法还原,原始数据和还原后数据完全一样. 这从表 2 中原始数据,水印数据和还原后数据的均值和方差变化也可以看出,同时水印数据和原始数据相比,可以看出其均值和方差变化的并不太大,这在一定程度上可以说明 3.2 节可逆水印算法进行水印操作后,对用户有一定的不可见性. 从表 2 可以看出,当嵌入比例增大时,原始数据和水印数据之间均值和方差的变化也变大,所以需要在嵌入信息量(与嵌入比例有关)和不可见性进行折衷考虑.

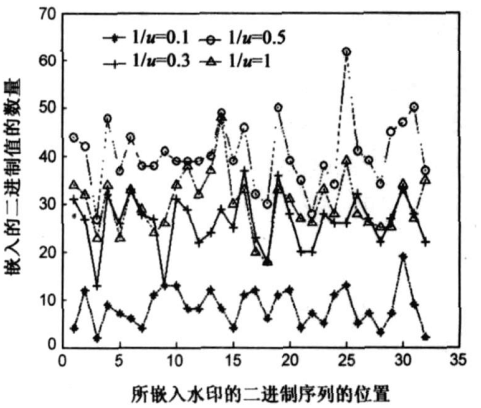

图 2 所嵌入水印二进制序列中每个值的嵌入次数

通过以上算法分析和实验分析,当用户输入错误的参数时,即使水印数据没有任何篡改也不能够还原出原始数据,图

4 所示为错误输入不同参数时原始数据和还原后的数据分布比较,图 4 在一定程度上可说明 3.2 节的可逆水印算法具有可行性.

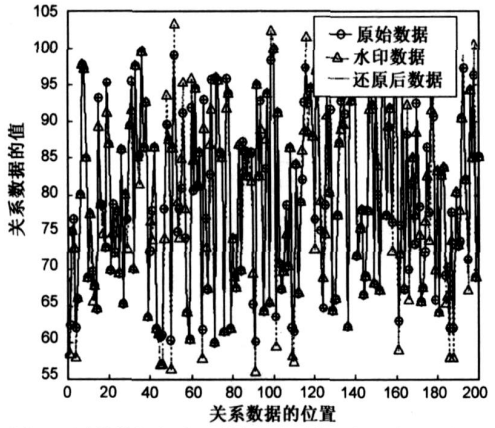

图 3　原始数据与水印数据及还原后数据分布比较

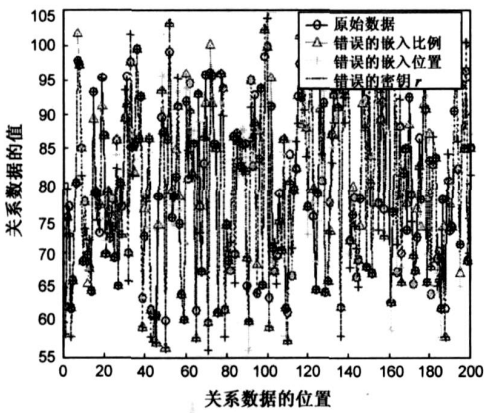

图 4　错误输入参数时原始数据和还原后数据分布比较

表 2　原始数据、水印数据和还原后数据的均值、方差变化

嵌入比例	原始数据		水印数据		还原后数据	
	均值	方差	均值	方差	均值	方差
0.1	80.2999	11.50457	80.3399	11.60966	80.2999	11.50457
0.3			80.2539	11.82022		
0.5			80.2659	11.98505		
1			79.9459	12.11516		

5　结论

关系数据库可逆水印技术的提出,为关系数据库数据的所有权保护和认证提供了一个新方法,本文只是抛砖引玉,还需在技术上作深入研究.将来这方面的工作将更多地集中于可逆水印方法研究,特别是可逆水印的盲测问题研究.

参考文献:

[1] Fridrich J, Goljan M, Du R. Invertible authentication watermark for JPEG images[A]. Proc of Information Technology: Coding and Computing[C]. Washington, DC: IEEE Computer Society Press, 2001. 223 - 227.

[2] C W Honsinger, P Jones, M Rabbani, J C Stoffel. Lossless Recovery of an Original Image Containing Embedded Data[P]. US Patent: 6278791, August 21, 2001.

[3] J Dittmann, S Katzenbeisser, C Schallhart, H Veith. Provably Secure Authentication of Digital Media Through Invertible Watermarks[EB/OL]. http://eprint.iacr.org/2004/293, Nov 7, 2004.

[4] J Fridrich, M Goljan, R Du. Invertible authentication[A]. Proceedings of the SPIE vol. 3971, Security and Watermarking of Multimedia Contents III[C]. Washington, DC: SPIE Press, 2001. 197 - 208.

[5] J Fridrich, M Goljan, R Du. Lossless data embedding-new paradigm in digital watermarking[J]. EURASIP Journal on Applied Signal Processing, 2002, (2): 185 - 196.

[6] Josep Domingo-Ferrer, Francesc Sebé Invertible spread-spectrum watermarking for image authentication and multilevel access to precision-critical watermarked images[A]. Proceedings of the International Conference on Information Technology: Coding and Computing[C]. Washington, DC: IEEE Computer Society Press, 2002. 152 - 157.

[7] D Maas, T Kalker, F M Willems. A code construction for recursive reversible data-hiding[A]. Proceedings of the ACM Workshop on Multimedia[C]. New York: ACM Press, 2002. 15 - 18.

[8] J Dittmann, O Benedens. Invertible authentication for 3d-meshes[A]. Proceedings of the SPIE vol. 5020, Security and Watermarking of Multimedia Contents V[C]. Washington, DC: SPIE Press, 2003. 653 - 664.

[9] M Steinebach, J Dittmann. Watermarking-based digital audio data authentication[J]. EURASIP Journal on Applied Signal Processing, 2003, (10): 1001 - 1015.

[10] Agrawal R, Kiernan J. Watermarking relational databases[A]. Proc 28th VLDB Conference[C]. Hong Kong: Springer-Verlag, 2002. 155 - 166.

[11] Sion R, Atallah M, Prabhakarl S. Watermarking Relational Databases[R]. Indiana: the Center for Education and Research in Information Assurance and Security of Purdue University, 2002.

[12] Zhi-hao Zhang, Xiao-ming Jin, Jian-min Wang, De-yi Li. Watermarking relational databases using image[A]. Proc of IEEE Conf on Machine Learning and Cybernetics[C]. New Jersey: IEEE Press, 2004. 1739 - 1744.

[13] Y Zhang, X N Niu, D N Zhao. A method of protecting relational databases copyright with cloud watermark[J]. International Journal of Information Technology, 2004, 1(4): 206 - 210.

[14] Zhang Y, Niu X M, Wu D, Zhao L, Li J C, Xu W J. A method of verifying relational databases ownership with image water-

mark[A]. Proceedings of the 6th International Symposium on Test and Measurement[C]. Beijing: Chinese Society for Modern Technical Equipment, 2005. 6316 - 6319.

[15] Y Zhang, X M Niu, A Khan, Q Li, Q Han. A novel method of watermarking relational databases using character string[A]. Proceedings of the IASTED International Conference on Artificial Intelligence and Applications 2006[C]. Calgary: ACTA Press, 2006. 120 - 124.

[16] Yingjiu Li, Huiping Guo, Sushil Jajodia. Tamper detection and localization for categorical data using fragile watermarks[A]. Proc ACM Workshop on Digital Rights Management (DRM)[C]. New York: ACM Press, 2004. 73 - 82.

作者简介:

张 勇 男,1976 年 12 月出生于江苏省泗洪县,现为哈尔滨工业大学深圳研究生院副教授,在国内外发表学术论文 19 篇,申请国家发明专利 2 项(其中已授权 1 项).

E-mail: zhangyong076@gmail.com

牛夏牧 男,1961 年 5 月出生于辽宁省锦州市,现为哈尔滨工业大学计算机学院教授,博士生导师,获国家科技进步奖、航天部科技进步奖、机械电子部科技进步奖、国家教委科技进步奖 6 项,2002 年全国百名优秀博士论文获得者,2004 年教育部"新世纪人才计划"获得者,在国内外发表学术论文 70 余篇.申请国家发明专利 4 项.

E-mail: xiamu.niu@hit.edu.cn