基于密码学的访问控制和加密安全数据库

袁 春[1], 文振 [2], 张基宏[2], 钟玉琢[1]
(1. 清华大学深圳研究生院信息学部, 广东深圳 518055; 2. 深圳大学信息工程学院, 广东深圳 518060)

摘 要: 在数据库安全领域的研究中,“数据库管理者”对数据库安全带来的安全隐患越来越受到研究者的关注, 尤其是在基于互联网的网络应用提供商(Internet service provider)模式的数据库系统中, 这种安全威胁更为严重. 传统的数据库访问控制(存取控制)方法, 对此安全隐患不能提供有效的安全防范. 密码学的安全数据库技术, 因为其基于数学难解问题的计算复杂性, 成为解决数据库安全问题的日渐重要的方法. 本文分析了迄今为止各种不同的, 针对分布式数据库应用的安全威胁, 并对密码学安全数据库中的基于密码学的访问控制算法和加密数据库技术进行了综述, 并对各类方法的密码学原理, 算法特性以及其优缺点进行了分析和描述.

关键词: 密码学; 安全数据库; 访问控制; 加密数据库
中图分类号: TP311.11 文献标识码: A 文章编号: 0372-2112 (2006) 11-2043-04

Progress of Cryptographic Access Control and Encryption Security Database

YUAN Chun[1], WEN Zhen-kun[2], ZHANG Ji-hong[2], ZHONG Yu-zhuo[1]
(1. *Graduate School at Shenzhen, Tsinghua University, Shenzhen, Guangdong* 518055, *China*;
2. *College of Information, Shenzhen University, Shenzhen, Guangdong* 518060, *China*)

Abstract: Researchers on security database concern more and more the security threats from DBA(database administrator), for its resulting in increasing crime to internet commercial databases which exist in the mode of ASP(Application service provider) recently, and its theoretic and technical hardness. Traditional access control could not provide enough security for the purpose, cryptographic security database becomes a promising scheme favored by its computing complexity of mathematical difficulties. We analyse various and up to date security threats to distributed database application systems, and give an overview for cryptographic access control and encryption database technology. Each class and approach is described and evaluated with its cryptographic principle, algorithm characters and positive and negative performance. We focus on the term of cryptanalysis of the schemes, which is considered the crucial points of the technology.

Key words: cryptographic; security database; access control; encryption database

1 引言

安全数据库技术的发展受一些密切联系的因素影响, 例如数据库技术本身, 数据库应用环境的逐渐扩大, 以及数据库面临的各种安全威胁. 研究者们提出了许多方法来设计安全数据库.

访问(存取)控制(Access control)可以看作最基本的安全数据库方法, 从数据库产生之日起, 该技术就被采用, 并随着数据库技术的发展而不断更新和进步. 访问控制算法可以归纳为三种类型: 自主型访问控制 discretionary access control (DAC)[1~4]、强制型访问控制 mandatory access control (MAC)[5~7] 和角色型访问控制 role based access control (RBAC). 自主型和强制型访问控制算法又被许多研究者们扩展成为面向对象的数据库访问控制类型. 访问控制方法一般是采用关系矩阵的形式来表示数据库的主体和其相对应的数据库客体对象的访问及权利关系.

1.1 数据库的安全威胁

今天, 人们处于互联网和“无所不在的计算”的时代, 各种数据库应用在互联网上被数据库服务提供商以在线的形式提供, 或用于电子商务, 或用于海量数据查询等各种应用. 在这些应用中, 数据库本身变得日益复杂, 同时, 为用户提供的服务响应也越来越丰富, 于是, 基于网络应用的数据库安全威胁也日益严重, 情况也变得日益复杂.

2004 年 3 月 17 日, 黑客侵入一家重要的消费者银行代理机构的高度机密数据记录, 1400 人的银行信用记录被窃取. 2002 年 1 月, 全球健康 Trax 公司的网络被入侵, 成千的银行

收稿日期: 2006-01-19; 修回日期: 2006-06-20
基金项目: 国家自然科学基金(No. 60573132); 国家自然科学基金(No. 60503063); 深圳市项目(No. 200507)

信用卡号被盗取.

面对这些严重的网络安全事件,网络数据库应用的安全漏洞成为研究者们必须面对的问题.数据库的安全威胁来自于许多不同的途径.如果我们跟踪一个分布式网络数据库应用的数据流过程,两种主要的安全问题就可以被归纳出:安全数据传输和安全数据存储及访问[8].如果把数据库的安全威胁分为物理的和逻辑的两个方面,那么在逻辑方面的威胁可以包括:信息的暴露,非法访问和篡改数据,以及服务拒绝等;在物理方面的威胁可以包括:访问密码的强行获取,窃取或破坏存储设备,电源破坏等.逻辑威胁和物理威胁都可以是故意的和意外的,如果安全威胁按照其所受攻击的来源来分类,可以分为外部侵入,内部管理漏洞和系统管理员.

1.2 密码学安全数据库

针对上述这些数据库安全威胁,传统的访问控制算法一直以来是主要的安全防范机制,但它已经不能满足今天日益复杂的安全威胁和日益丰富的私有性保护需求.例如,在访问控制模式下,数据库管理员拥有最高的特权,这就有可能产生安全漏洞.为了克服这些包括数据库管理员特权带来的安全漏洞,密码学技术被采用并深入到安全数据库技术的各个方面,可以统称为基于密码学的安全数据库技术.

根据统计,大多数网络数据库的数据被盗事件是源自数据库系统的内部人员,包括数据库操作者和管理者.所以,限制数据库管理员特权滥用,以保护数据的安全隐私就成为当今重要的研究课题.最初的方法是将密码学和访问控制结合起来[8~10],如安全管理方法 SA(security administer)[10]和安全字典方法 SD(security dictionary)[9].两种方法都是试图减少数据库管理员对安全相关的数据操作的干涉.

为了更有效地消除 DBA 安全漏洞,加密数据库的方法被提出,并且在安全数据库技术中崭露头角.然而,这项技术又带来了一个问题,就是对于加密的数据库内容,很难直接完成数据库的关系逻辑操作和运算.为克服这个障碍,许多加密数据技术被提出[11~14],如利用私有同态的原理,该原理是指对加密的数据进行一定的运算;利用次序保持加密数据库算法[15,16],该算法的优势在于加密数据库的 B 树索引可以继续保持有效,以用于加密数据库的查询操作.

需要指出的是,对安全数据库足够安全的需求和对加密数据库方便的查询能力的需求,二者不可避免地存在着矛盾.这就要求我们在设计一个密码学安全数据库的时候,必须对二者进行仔细的考察和折衷以选择合适的密码学方法.也就是说,必须认真衡量数据库的安全特性在具体应用中安全威胁下带来的利益和它的查询方便性的损失.

2 密码学访问控制

在传统的数据库访问控制算法中存在一些固有的缺陷.首先,它是基于服务器方式下的安全管理,DBA 拥有最高的特权,从而产生内部安全漏洞;其次,访问控制是基于引用监督策略和访问控制矩阵,这样,在分布式网络数据库系统中,来自于网络的攻击有可能采用绕过访问控制环节,故意提升用户等级或者修改相关系统文件等方式进行攻击.针对这些安全漏洞,研究者们设计出基于密码学的安全访问控制算法.

丹麦科技大学的研究组提出了一个针对分布式文件系统密码学安全访问控制算法(2003)[8].该算法将加密算法和访问控制算法相结合来提高传统访问控制算法的机密性,同时提供完整性保护.算法主要基于一个开放的网络体系,传统的访问控制算法不能提供可靠的安全保障.在算法中,客户被分别赋予针对系统中存储的数据"读"和"写"两个权利.在读的状态下,客户被授予"对称密钥",从而对从服务器获取的数据进行解密运算,还被授予"解密密钥"来验证加密数据的签名信息从而进行认证和完整性检验.在写的状态下,用户被分配"对称密钥",对即将写入并发送到服务器的数据进行加密运算,同时被分配"加密密钥"对数据进行消息摘要的完整性计算及产生签名.服务器端被赋予"解密密钥"来检验数据的完整性.该算法虽然能限制服务器端的部分安全漏洞,但是算法的计算负责度比较高,加密密钥和解密密钥一般为公钥体系结构,而且系统的存储负荷也比较大.

IBM Watson 实验室的 Jingmin He 和 Min Wang 提出了一种基于密码学的关系数据库管理系统(1998)[9].该方法引入了一种安全字典的概念,它可以为数据库提供许多安全服务功能,其中包括安全用户管理机制来适应各种关系数据库管理系统的安全需要.在该方法中,需要一个安全环境来存储安全字典.安全字典包括许多索引表和视图,这个安全字典有数据库服务器来维护,并只能通过系统命令的方式进行更新.对它的访问是有一个严格的授权和认证策略来控制的.该系统尚未提出有效的方案来实现在数据库服务器中建立和维护一个安全操作环境(SOE),尤其是当安全操作环境中的数据需要被服务器频繁更新的情况下.

3 加密数据库

加密数据库在安全数据库中越来越受到关注,尤其是在互联网中日益增多的"应用服务提供商"模式下,数据库也成为一种网络服务模式.当数据库的拥有者将其提供给网络应用提供商,以构成更广范围的电子商务服务,这就会带来两种重要的安全需求:一是如何保护这些数据库内容或数据的机密性或隐私,以防止来自互联网的外部黑客的攻击,同时也是保护内容提供商的知识产权;二是如何防止网络应用提供商(运营商)滥用数据资源或由此而引起的其他安全漏洞.一种直接的方法就是建立安全加密数据库,在网络服务提供商(运营商)处,数据是无法被解密的,但是对于合法的用户,可以被授予相应的解密权利来获取数据内容.

加密数据库就意味着数据库的管理系统能够对加密的关系数据进行相关操作.但是,在密码学的理论中,对于加密的数据进行关系数据库的关系运算,还没有可行的解决方案.于是研究者们求助于一种私有同态技术(privacy homomorphisms),它可以对加密的数据进行一些基本的运算.这样加密数据库的管理系统就可以利用 SQL 语句来对加密的数据库数据进行一些基本的操作,如逻辑比较,基本运算或格式匹配等.从而在服务提供商处不会产生信息泄漏的问题.

3.1 查询加密数据库

在实现加密数据库时,仍然会存在一些具体的问题.一是

如何加密数据内部的关系数据,也就是如何确定加密的层次,如磁盘的页,整个关系表,单个记录或者是某几个属性字段.采取任何一种方法都会有利有弊.二是如何将 SQL 应用于加密数据的逻辑运算或操作,尤其是实现加密数据的聚合,区间查找和格式匹配等逻辑操作.

Davida 等人提出了一种数据库加密系统(1981)[11],这种方法是面向记录加密,对整个记录采用分组密码的方式进行加密,这样就可以防范针对某个属性字段的安全攻击.整个加密后数据块在解密时,每个属性字段可以被单独解密获得,而不会也不必解密其他的属性字段.这个方法可以很好的解决那些数据记录需要整体加密,而各个字段可以分别被不同权利的访问者解密使用的数据库应用场景.中国剩余定理是该方法的技术原理,问题是,因为中国剩余定理是一种通用素数的运算,如何选择素数,以及如何构建一种共用的方法,都是非常复杂的计算过程,所以该方法具有较困难的扩展性.

Hacigumus 等人提出了一种安全数据库服务提供模型,它通过改进的 SQL 对加密的数据库进行关系运算(2002)[12].在该方法中,关系表中的记录采用常规的分组加密方法进行加密,对于每个属性字段,需要增加一个附属字段来标示该属性字段(未加密前)所属的数值区间.该附加字段的值用于加密数据库的查询操作,这样,会出现多个加密记录的相同字段对应相同的附加字段值,在进行某字段的数值匹配查询时,就会产生多个匹配记录,然后,一个"后处理"的过程来解密这些记录,从而最终获得准确的查询结果."后处理"的计算复杂度取决于属性字段的分段机制.该方法考虑了数据库技术的许多相关方面来适应加密数据库查询的特殊要求,例如语法规则,代数架构和查询优化等.它的问题在于,如果字段值的分段过于粗糙,"后处理"的计算复杂度就越高,如果字段值的分段越细,加密数据的安全性就越低.

Song 等人提出了一种实际的技术手段来解决加密数据库的查询问题(2000)[13],该方法针对的环境是,在一个"不可信"的服务器中存储了一些加密的文档,用户需要通过查询包含某个字(关键词)的文档.该算法把流密码和分组密码结合起来,以加速和减轻算法的处理开销.具体方法是,文档被分割成字的序列,每个字被一段伪随机序列以特殊的结构进行加密,当一个用户需要在加密的文档中查询一个字时,他给不可信服务器发送一段有关该字的"最小信息",然后服务器进行查询并返回查出的结果.共有四种方案被提出,以逐步增加查询的安全等级.方法中,服务器可以获得用户需要查询的关键词的明文,同时服务器还可以获取关于这个词的加密值的一些信息,所以算法的安全性是有待提高的.

3.2　次序保留的加密数据库

一般来说,加密的数据库在执行查询操作时,索引文件是无法支持的,由于索引文件的失效,还会引申带来其他一些不利因素,即数据库在非加密时的具有的查询便利特性,在加密数据库中就很难支持了.为了解决这一问题,一种次序保留的加密数据库技术被提出了.

然而,需要指出的是,对于加密数据库的高安全性需求,是和针对加密数据库中加密数据查询操作的便利性相矛盾的.一般来说,如果数据库因为加密带来的性能降低的程度是可控的,用户更愿意使用加密的数据库而不是"非加密"的数据库.

Ozsoyoglu 等人提出了一种可查询的加密数据方法(2003)[14].该方法采用一组严格单调增加的,而且是可逆的多项式函数,作为加密运算,以一种多层嵌套的方式来构成次序保持的加密数据库,这样,数据库原有的 B+ 树等类型的索引文件可以仍然保持有效,从而获得高效的加密数据库的查询.该算法的主要贡献在于其无损的多层加密函数,把对数运算和多项式运算结合起来,以防止整数的溢出错误和实数的精度错误.该算法对浮点数的精度(有效值)部分和和指数部分分开进行运算.该方法的主要问题在于没有考虑数据库字段的输入数据的分布特性,也就是说加密后的数据和原数据的概率分布特性很相似.该方法需要一个独立的安全第三方来进行加解密运算.

另一种称为 OPES(order preserving encryption scheme)次序保持加密方法[15],是基于如下的应用环境条件:(1)数据库软件的存储系统是安全性脆弱的;(2)数据库软件的安全可信的;(3)所有的磁盘数据是加密的,包括数据库关系表,属性字段名和相应的值.该方法采用数值分布变换的方式来实现次序保留,主要包括三个步骤:(1)建模:对输入和输出的目标分布函数进行建模,分别变换成分段线性函数;(2)均匀化:明文数据库被变换成一个平滑数据分布的数据库,数据库中的值是均匀分布的;(3)变换:平滑数据库被变换成一个密文数据库,使数据库的值成为目标分布的特性.该方法的安全特性是基于一个平滑函数,来作为加解密函数,函数中的两个参数作为密钥.

该方法的缺陷在于:只能防范密文攻击;算法成立的前提条件是攻击者不掌握原始数据的任何相关信息(数值分布特点),这是个很难满足的条件.另外,此方法不能很好适应数据库在服务提供模式下的应用安全,因为方法没能对数据库的管理者进行有效的安全限制.

4　结论

基于密码学的安全数据库技术是一个非常广阔的研究领域,它和数据库技术本身的发展密切相关,同时也和数据库网络服务模型的不断丰富密切相关.尤其现在,应用服务提供商将数据库技术广泛应用于商业、金融、政务等领域,同时也带来了越来越多的数据库安全隐患,例如网络数据库的内容滥用,基于互联网的数据库网络入侵,商业数据库的盗版问题以及数据库管理方面的安全漏洞等.

人们提出了许多基于密码学的安全数据库技术,包括密码学访问控制技术,加密数据库技术和安全 SQL 数据库技术,基于客户端的安全数据库技术等.尽管加密数据库被认为是安全数据库非常有前途的技术,但对于该项技术的具体应用还需要进行更深入的探讨和研究.

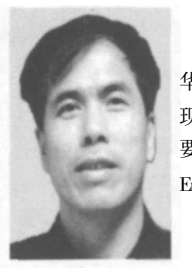
参考文献:

[1] Lampson B W. Protection [A]. Proc. of The 5th Princeton Symp

posium on Information Science and Systems [C]. Princeton, New Jersey, USA, March, 1971. 437- 443.
[2] Denning P J. Protectionr principles and practice [A]. Proc. of the Spring Joint Computer Conference [C]. Montvale, New York, 1972. 40.
[3] Harrison M A, et al. Protection in operation system[J]. Communications of the ACM, 1976, 19(8): 14- 24.
[4] Denning D E. Database Security [M]. Annual Review Inc, 1988.
[5] Costich O, et al. A multilevel transaction problem for multilevel secure database systems and its solution for replicated architecture [A]. Proc. of IEEE Computer Society Symposium on Research in Security and Privacy [C]. Oakland, CA, 1992. 192- 203.
[6] Denning D E, et al. A multilevel relational data model [A]. Proceedings of the First International Workshop on Object-Oriented Database Systems [C]. Pacific Grove, 1986.
[7] Dwyer P A, et al. Multilevel security in database management systems [J]. Computers and Security, 1987, 6(3): 252- 260.
[8] Harrington A, et al. Cryptographic access control in a distributed file system [A]. Proceedings of the Eighth ACM Symposium on Access Control Models and Technologies [C]. Como, Italy, 2003. 158- 165.
[9] He J, et al. Cryptography and relational database management system [A] Proc. of Database, Engineering and Application Symposium [C]. Grenoble, France, 2001. 273- 284.
[10] Mattsson U, et al. Secure Data Functional Overview [M]. Protegity Technical Paper TWP-0011, 2000.
[11] Davida G L, et al. A database encryption system with subkeys [J]. ACM Transactions on Database Systems, 1981, 6(2): 312 - 328.
[12] Hacigumus H, et al. Executing SQL over encrypted data in the database-service-provider model [A]. ACM SIGMOD Conference [C]. Madison, Wisconsin, USA, 2002. 216- 227.
[13] Song D X, et al. Practical techniques for searches on encrypted data [A]. IEEE Symposium on Security and Privacy [C]. Los Alamitos: IEEE Computer Society Press, 2000. 44- 55.
[14] Gultekin S C, et al. Anti-tamper databases: Querying encrypted databases [A]. Proc. of the 17th Annual IFIP WG 11.3 Working Conference on Database and Applications Security [C]. Estes Park, Colorado, August. 2003. 133- 146.
[15] Agrawal R, et al. Order preserving encryption for numeric data [A]. Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data [C]. Paris, France, 2004. 563- 574.

作者简介:

袁 春 男, 1969 年出生, 清华大学深圳研究生院副研究员, 1999 年, 2003 年分别获得清华大学计算机系硕士、博士, 2003 至 2005 年在法国国家研究院安全媒体信息系统实验室任博士后、研究员, 研究方向: 安全多媒体技术, 数据安全, 安全数据库, 多媒体家庭等.



文振 男, 1962 年 3 月生于广东信宜, 清华大学计算机科学与技术专业毕业, 工学硕士, 现为深圳大学信息工程学院计算机系副教授, 主要研究方向为流媒体编码、基于内容视频检索. Email: wenzk@szu.edu.cn

张基宏 男, 1964 年 6 生于江苏海安, 东南大学无线电专业毕业, 博士学位, 现为深圳大学教授, 主要研究方向有图像编码、矢量量化和模糊逻辑.

钟玉琢 男, 1938 年出生, 清华大学深圳研究生院信息学部主任, 教授, 博导, 主要研究方向: 多媒体计算机技术, 视频压缩, 视频点播, 视频网络流化和多媒体家庭平台.