几种数据库加密方法的研究与比较

赵晓峰,叶 震
(合肥工业大学 计算机与信息学院,安徽 合肥 230009)

摘 要:随着数据库技术在日常经济生活中应用的不断增加,数据库安全日益成为人们关注的热点。而目前数据库的安全性主要通过访问控制来保障,但是当访问控制被攻破时整个数据库的安全体系也随之瓦解。目前解决此问题的主要方法是采用数据库加密。文中对目前主要的几种数据库加密方法进行了探讨和研究,并指出了数据库加密的发展方向。

关键词:数据库加密;秘密同态;子密钥;密钥管理;密文索引
中图分类号:TP309.7 文献标识码:A 文章编号:1673-629X(2007)02-0219-04

Research and Comparison of Several Database Encryption Technologies

ZHAO Xiao-feng, YE Zhen
(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

Abstract: With the development of database usage into economy and production of our daily life, the database security is more and more concerned. Nowadays the main technology used in the database security is access control, but when the access control mechanics is broken through, the whole security system is collapsed. Database encryption is introduced to solve this problem. Several technologies of database encryption are discussed in this paper, as well as the direction of corresponding researches.

Key words: database encryption; privacy homomorphism; subkey; key management; ciphertext index

0 引 言

数据库技术产生于 20 世纪 60 年代末,是信息系统的核心技术。随着经济的日益增长,数据(仓)库系统被越来越多地应用于各种机构之中。而很多机构又将其核心的机密信息存于数据(仓)库之中。所以,数据库的安全性就显得越来越突出。而数据库加密技术也逐渐成为关于数据库安全的研究热点之一。

1 研究数据库安全技术的必要性

目前市场上的大部分数据库系统都符合 C1 和 C2 级安全标准,而美国也拥有符合 B1 级标准的军用版 Oracle 系统,中国虽然也拥有一些符合 B1 级标准的安全 DBMS 原型,比如:OpenBASE,Cobase,以及 2003 年中科院信息安全国家重点实验室基于开放源代码的数据库管理系统 Postgre SQL,开发出安全数据库系统 LOIS。总体来说,与国外主流数据库产品相比,这些研究成果在安全性和可用性上还有一定的差距。开发一个具有自主知识产权的能够符合实际应用要求的安全系统还是一个尚待解决的问题。

根据 2004 年底的统计,几大国外数据库管理系统在国内的市场占有率达到 95%,国产数据库的总市场容量大约为 3.5%,其他已开源的产品大约占 1.5%。国外的数据库产品不提供源程序代码,也很少有可供公开调用的内核接口,这些都加大了自主安全保护的技术难度。加之发达国家限制 C2 级以上安全级别的信息技术与产品对中国的出口,故研究开发数据库安全控制技术具有重要的现实意义。

2 目前数据库加密所存在的问题

目前的数据库系统主要采用访问控制的方式保证系统的安全性,但是当访问控制被攻破时整个数据库的安全体系也随之瓦解。目前解决此问题的主要方法是采用加密数据库。但是,多年来,对数据库的加密一直是一个未能完全解决的难题。目前对于数据库加密体系主要存在以下两个问题:

(1)由于数据库系统要执行大量的查询操作,这就要求数据库加、解密算法既要保证系统的安全性,又要

收稿日期:2006-04-26
作者简介:赵晓峰(1978-),男,浙江绍兴人,硕士研究生,研究方向为计算机网络与信息安全;叶 震,副研究员,研究方向为信息安全。

保证对密文数据库查询的方便和快速性。但是目前已提出的数据库加、解密算法,及在其之上的查询、添加、删除方法还都不能同时满足系统的安全性和易用性要求。而寻找一个在不影响用户正常使用(如:添加、删除、查询)前提下能快速对数据进行加解密的方法也成为数据库加密体系研究中迫切需要解决的问题之一。

(2)目前针对数据库加密体系中密钥管理的应用和研究大多是以“可信第三方”为基础的。这种基于“可信第三方”的密钥管理方法,认为存在一个可信任的第三方(如:PKI 体系中的 CA,DBMS 中的 DBA 等),并将用户密钥交给他保存。“可信第三方”拥有获取和更改用户密钥的权利。这种方法相对来说比较灵活,比如:当用户忘记了自己的密钥,可以由“可信第三方”帮其恢复,并且对于抵御来自外部的攻击是非常有效的。但是,这种密钥管理方法与密码学中用户本人以外的任何人都不可信的理念相悖。也就是说整个加密系统的安全性是建立在“可信第三方”的可信度基础上的,一旦可信第三方变为“不可信”,整个加密系统的安全性将会受到很大威胁。研究表明目前数据库安全的主要威胁是来自于系统内部而不是外部。但是遗憾的是目前还没有提出基于“不可信第三方”的安全且灵活的密钥管理方案。

3 加密数据库要考虑的几个问题

3.1 在什么层次上对数据库进行加密

可以考虑在 3 个不同层次实现对数据库数据的加密,这 3 个层次分别是 OS, DBMS 内核层和 DBMS 外层。

在 OS 层,从操作系统的角度来看 OS 层位于 DBMS 层之下,所以无法辨认数据库文件中的数据关系,也就无法合理地产生、管理和使用密钥,因此,在 OS 层对数据库文件进行加密,对于大型数据库来说,目前还难以实现。

在 DBMS 内核层实现加密,是指数据在物理存取之前完成加/解密工作。这种方式的优点是加密功能强,并且加密功能几乎不会影响 DBMS 的功能;缺点是在服务器端进行加(解)密运算,加重了数据库服务器的负载,并且因为加(解)密是在 DBMS 内核中完成,就势必需要数据库供应商对其进行技术支持,这一点不容易实现。

另一种做法是将数据库加密系统做成 DBMS 的一个外层工具。采用这种加密方法的优点是可扩充性强,数据库的加解密系统可以做成一个独立于 DBMS 的平台,不需要数据库供应商进行技术支持,并且可以将加密密文直接在网上传输;缺点是数据库的功能和查询效率会受一些限制。

3.2 加密单元的选择

可以考虑以字段、记录和数据表为最小加密单元。

按记录加密存在的主要问题是加密后的数据很难再写入原来的数据库中。

按字段加密应该是最好的方式,因为这样会使要加密数据的长度最小。但是这种方式需要考虑数据转换的问题。具体选取哪种方式要看现实应用的需要。

4 数据库加密系统应满足的要求

由于数据库具有数据复杂、数据的查询操作非常频繁且数据存储时限相对较长等特点,所以应用于数据库的加、解密算法及相应的密钥管理机制应满足以下要求:

(1)数据库加密系统应满足的首要条件是保证数据的安全性。在此方面要求加密算法保证数据的保密性和完整性,防止未授权的数据访问和修改。

(2)数据库中存在大量的查询操作,因此加解密效率要求较高,不能引起数据库系统的性能大幅度下降。

(3)数据库组织结构对于数据库管理系统而言不能有太大的变动,应尽可能做到明文和密文长度相等或至少相当。

(4)由于时限较长和密钥的复杂,密钥管理机制应更加安全、灵活和坚固。

5 目前数据库加密的主要方法及研究现状

5.1 主要方法

目前对于数据库加密的研究主要集中于高效数据库加/脱密引擎的寻找,以及基于此的快速查询、插入、删除方法上。关于这个方向主要的方法有以下几种:

1)秘密同态技术。

寻找一种既保证数据库安全性,又保证使用(如:查询,插入,删除)的方便性的加密方法一直是数据库加密的主要研究方向。为了提高密文数据库的查询效率,Rivest 等人在文献[1]中提出秘密同态的概念。

定义 1(秘密同态) 假设 E_{k1} 和 D_{k2} 分别代表加密、解密函数,明文数据空间中的元素是有限集合 $\{M_1,\cdots,M_n\}$,α 和 β 代表运算,若 $\alpha(E_{k1}(M_1),\cdots,E_{k1}(M_n)) = E_{k1}(\beta(M_1,\cdots,M_n))$ 成立,则称函数族 $(E_{k1},D_{k2},\alpha,\beta)$ 为一个秘密同态。

文献[1]中提出的秘密同态技术能够对数据直接在密文的状况下进行操作,从而有效提高对密文数据库的查询速度。但是该方法对已知明文攻击存在一定安全隐患,所以 Domingo 等人在文献[2]中对其进行了

改进。此后国内外也有很多基于秘密同态技术的研究。秘密同态技术能够对未经解密的密文数据进行查询,大大提高了密文数据库的查询效率。但是,因为该方法对加密算法提出了一定的约束条件,使得满足密文同态的加密算法的应用不具有普遍性[3]。

2)密文索引技术。

提高密文数据库查询效率的另一种主要方法是密文索引技术。假设属性 A 是用户的查询属性,为 A 建立索引 A',A 是对用户保密的,用户只能看到其索引 A',这样既保证了用户查询的方便性又保证了敏感数据的安全性。对于加密粒度为字段级和记录级的加密方法比较适合于建立索引,而对于加密粒度为属性列的加密算法,因为算法是以属性列为最小加密单元的,即使为属性列建立索引在检索时也需要对整个属性列进行解密,所以并不适合于建立密文索引机制。

目前有许多关于密文索引的方法。文献[6]提出了一种基于哈希算法的分散密文索引技术,假设敌手只拥有密文数据及其对应的索引,由于索引中的数据地址是以密文形式存放的,找不出密文与索引的对应关系,该方法能对抗敌手的静态分析,但是对于动态跟踪技术存在一定的安全隐患。文献[4]提出了一种可以防范内部攻击的基于元组的索引方法,但是对每一个索引的查询都会返回所有相匹配的元组,降低了查询效率。另外还有顺序索引技术、数组索引技术和矩阵索引技术等也都针对不同问题被先后提出。对于大型数据库建立索引,由于数据量较大,建立的索引不能完全存放于主存之中,目前大多采用 B+ 树的方法将索引存于外存之中,当应用时再将部分索引调入主存。采用 B+ 树存储存在一定缺陷,因为数据库需要进行大量的存储和删除操作,随着节点的增删,势必会引起 B+ 树的不平衡,从而影响查询效率。所以,对于小型数据库有研究提出以矩阵的形式存放索引表,并将索引表一次性存入主存之中,以提高查询效率。目前,大多数的密文索引技术都是针对于外部攻击的,虽然也有一些针对于内部攻击的密文索引技术,但是在安全性和易用性上还存在一定问题[5]。

3)子密钥加密方法。

传统的基于记录的数据库加密的方法存在一个问题,因为数据是以记录为单位进行加密的,所以在查询时需要对整个字段进行解密(或对需查询的明文进行加密)以后再进行查询,这就必然增加了查询开销。为了解决基于记录的数据库加密技术存在的问题,G. I. David 等人在文献[7]中提出了子密钥数据库加密技术。Hwang M-S 等人提出采用多级子密钥的安全模型以提高子密钥系统的灵活性和安全性[8]。子密钥加密算法的核心思想是根据数据库(特别是关系型数据库)中数据组织的特点,在加密时以记录为单位进行加密操作。而在解密时以字段为单位进行解密操作,系统中存在两种密钥,一种是对记录加密的加密密钥,另一种是对字段进行解密的解密密钥[9]。

子密钥加密方法,从一定程度上解决了针对记录加密方法的缺陷。但是,因为系统要保存两种密钥,这就增加了密钥管理的复杂性。这一点也是子密钥加密所急需解决的问题。

5.2　密钥管理

一个加密数据库由三部分组成:密钥、密文、明文,所以数据库的安全不仅与加、解密算法有关,也直接与密钥的安全性相关。一个好的数据库加密体系的必要条件之一就是拥有一套安全易用的密钥管理机制。所以对密钥管理的研究也就成为了数据库加密研究的一个重要方向[10,11]。

目前用于数据库加密的成熟的密钥管理(分发)方法(如:密钥转换表,PKI,Kerberos,等)大都是基于"可信第三方"的。而目前针对密钥管理的研究也多以"可信第三方"为基础。这种基于"可信第三方"的密钥管理方法相对更加灵活,并且对于抵御来自外部的攻击是非常有效的。但是由于用户的密钥可以被"第三方"(如:PKI 体系中的 CA,DBMS 中的 DBA 等)获取,这就使整个系统的安全建立在"可信第三方"的可信度之上。可以说基于"可信第三方"的密钥管理体系从某种意义上说存在着一定的安全隐患。但是,遗憾的是目前关于"非可信第三方"密钥管理体系的研究还相对很少。

5.3　关于数据库加密的其他研究方向

与数据库加密相关的还包括以下的研究方向:数据完整性研究,硬件加密方法的研究等。目前有很多研究将数字水印的方法用于数据库完整性之中。目前关于数据库水印的成熟算法还很少。在已有的研究中较著名的有 IBM Almaden 研究中心的 R. Agrawal 等人对数据库进行的水印的嵌入和攻击试验。该试验针对一个特定的数据库,在其中只包含数值型数据,且假定每个字段都能够添加水印,然后依据水印密钥和关键字确定需标记的字段及位置。美国 Purdue 大学的 R. Sion 等人提出采用"均方差"的特性对关系数据集合和数据库数据的数值型字段添加水印[12,13]。

通过研究发现,目前所提出的数据库水印技术大多是针对关系数据库的数值型数据进行加密的,并且对数据库进行了一定的约束。而现实中的数据库中的数据是多种多样的,所以当前提出的数据库水印技术还只是停留在初级研究阶段。

6 总 结

数据库加密是继访问控制后关于数据库安全的一个重要研究方向。采用数据库加密技术,即使系统的访问控制体系被黑客攻破也可以在一定程度上保证敏感信息不会外泄。文中指出了目前数据库加密技术存在的几点不足,并详细介绍了目前数据库加密的几种常用方法,及每种方法的优缺点。

参考文献:

[1] Rivest R L, Adleman L, Dertouzos M L. On Data Banks and Privacy Homomorphism[C]//In DeMillo R D. Foundations of Secure Computations. [s. l.]: Academic Press, 1978: 169 - 177.

[2] Domingo - Ferrer J. A New Privacy and Homomorphism Applications[J]. Information Processing Letters, 1996, 60(5): 277 - 282.

[3] 杨 勇,方 勇,周安民. 秘密同态技术研究及其算法实现[J]. 计算机工程, 2005, 31(2): 157 - 159.

[4] Hacigumus H, Iyer B, Li C, et al. Executing SQL over Encrypted Data in the Database - Dervice - Provider Model [C]//In Proceedings SIGMOD 02, Madison, International Conference on Management of Data. Wisconsin, USA: [s. n.], 2002: 216 - 227.

[5] Fischmann M, Gunther O. Privacy Tradeoffs in Database Service Architectures [C]//In Proceedings of BIZSEC 03, the First ACM Workshop on Business Driven Security Engineering. Berlin, Germany: [s. n.], 2003.

[6] 戴一奇,尚 杰,苏中民. 密文数据库的快速检索[J]. 清华大学学报, 1997, 37(4): 24 - 27.

[7] Davida G I, Wells D H, Kam J H. A Database Encryption System with Subkeys[J]. ACM Trans on Database Systems, 1981 (6): 31 - 37.

[8] Hwang M - S, Yang W - P. Multilevel Secure Database Encryption with Subkeys[J]. Data and Knowledge Engineering, 1997, 22(1): 117 - 131.

[9] 王庆梅,吴克力,刘凤玉,等. 一种子密钥数据库加密算法及其密钥管理方案研究[J]. 计算机工程与应用, 2003, 39 (11): 52 - 54.

[10] 戴一奇,尚 杰,陈 卫. 一种新的数据库加密密钥管理方案[J]. 清华大学学报:自然科学版, 1995, 35(4): 43 - 47.

[11] 余祥宣,崔永泉,崔国华. 分布式环境下数据库加密密钥管理方案[J]. 华中科技大学学报:自然科学版, 2002, 30(4): 43 - 45.

[12] Agrawal R, Kiernan J. Watermarking Relational Databases [C]//In 28th Int'l Conference on Very Large Database. Hong Kong: [s. n.], 2002.

[13] Damiani E, Vimercati D S C, Finetti M, et al. Implementation of a Storage Mechanism for Untrusted DBMSs [C]//In Proc of the Second International IEEE Security in Storage Workshop. Washington DC, USA: [s. n.], 2003.

(上接第 218 页)

经解决了该问题[1,6],此处不再介绍。

```
E:\论文\vpn\vpn1>build -cz
BUILD: Object root set to: ==> objfre
BUILD: /i switch ignored
BUILD: Compile and Link for i386
BUILD: Compiling and linking e:\论文\vpn\vpn1
Compiling - vpn.rc for i386
Compiling - revpn.c for i386
Compiling - protocol.c for i386
Compiling - miniport.c for i386
Compiling - receive.c for i386
Compiling - send.c for i386
Compiling - logprint.c for i386
Compiling - packetinfo.c for i386
Compiling - vpn.c for i386
Compiling - wdmdriver.c for i386
Compiling - passthru.c for i386
Compiling - des.c for i386
Linking Executable - objfre\i386\passthru.sys for i386
BUILD: Done

    12 files compiled
    1 executable built - 25 Warnings
```

图 3 IMD 编译结果

```
0.00060203 * * * e:\论文\vpn\vpn1\receive.c(186) * * *
0.00060622 = = = > VPN - PtReceiveRePacket...
0.00065595 * * * e:\论文\vpn\vpn1\receive.c(293) * * *
0.00066070 < = = = VPN - PtReceiveRePacket0.00066573 * * * e:
\论文\vpn\vpn1\receive.c(347) * * * 0.00066964 < = = =
VPN - PtReceive
0.00067523 * * * e:\论文\vpn\vpn1\receive.c(373) * * *
0.00067970 = = = > VPN - PtReceiveComplete
0.00068528 * * * e:\论文\vpn\vpn1\receive.c(407) * * *
0.00068947 < = = = VPN - PtReceiveComplete
0.01641913 * * * e:\论文\vpn\vpn1\miniport.c(174) * * *
```

图 4 IMD 输出结果

参考文献:

[1] Microsoft Corporation. Windows 2000 DDK [M]. US: Microsoft Corporation, 1985 - 2000.

[2] 朱雁辉. WINDOWS 防火墙与网络封包截获技术[M]. 北京:电子工业出版社, 2002: 210 - 228.

[3] 李 凌. Winsock2 网络编程实用教程[M]. 北京:清华大学出版社, 2003: 30 - 38.

[4] 蔡自兴,徐光祐. 人工智能及其应用[M]. 第 4 版. 北京:清华大学出版社, 2004: 311 - 316.

[5] Slattery T, Burton B. Advanced IP Routing in Cisco Networks [M]. Second Edition. Beijing: China Machine Press, 2001: 7 - 25.

[6] Microsoft Corporation. Windows XP DDK [M]. US: Microsoft Corporation, 2001.