维普资讯 http://www.cqvip.com

外包数据库系统安全机制研究*)

朱 勤[1,2] 于守健[1] 乐嘉锦[1] 骆轶姝[1]

(东华大学信息科学与技术学院 上海 200051)[1] (南通大学计算机科学与技术学院 南通 226019)[2]

摘 要 在外包数据库系统模式下,组织将数据库业务外包给外部数据库服务器运行,外包服务提供者为数据拥有者及数据库用户提供远程的数据库创建、存储、更新与查询服务。由于外包服务器本身并非完全可信,外包数据库系统带来了一系列安全问题。本文讨论了数据库安全体系的分类,探讨了外包数据库的基本结构,综述了外包数据库在数据加密、密文查询策略、隐私保护、数据完整性验证及基于数据库水印技术的数据库版权保护等安全机制方面的研究现状,介绍了其研究进展,并展望了安全外包数据库系统的发展方向。研究认为,整合多种安全机制,兼顾安全性和可用性两方面的合理需求,是构建外包数据库系统的技术关键。

关键词 外包数据库,数据库安全,数据库加密,数据完整性,数据库水印

Research on Security Mechanisms of Outsourced Database

ZHU Qin[1,2] YU Shou-Jian[1] LE Jia-Jin[1] LUO Yi-Shu[1]

(College of Information Science and Technology, Donghua University, Shanghai 200051)[1]

(College of Computer Science and Technology, Nantong University, Nantong, Jiangsu 226019)[2]

Abstract In the outsourced database model, organizations outsource their data management needs to an external service provider. The service provider offers mechanisms to create, store, update and query the database for the organizations. Since a server is almost never fully trusted, this model introduces several research issues related to data security. This paper discusses the classification of database security system, studies the architecture of outsourced database, overviews the state-of-the-art in database encryption, query policy of encrypted database, privacy protection, data integrity verification, and database copyright protection based on database watermarking, introduces the latest progress in security mechanisms of outsourced database, and looks ahead its future development. A conclusion is drawn at the end of this paper that it is critical to integrate various security mechanisms and trade off the security and the availability to build an outsourced database.

Keywords Outsourced database, Database security, Database encryption, Data integrity, Database watermarking

信息系统的外包服务,是指组织将其信息系统的开发、管理或维护的工作外包给专业性外部组织来完成的一种运行模式。随着信息技术在组织中发挥的作用日益增加,信息系统规模越来越大,涉及的技术也越来越复杂,外包服务把自己不擅长的东西(非核心业务)交给专业的外部组织去做,将主要精力集中于核心业务,从而优化资源配置,降低软硬件投资成本,提高组织机能和效率,充分发挥自身核心竞争力并增强自身应变能力。而日益高速和完善的计算机网络环境为这种服务提供了可能。

2002年,美国加州大学 Irvine 分校 Hakan Hacigumus 的论文"Providing Database as a Service"[9]首次提出了"作为服务的数据库"的概念,亦即外包数据库(Outsourced Database),并基于 Internet 构建了一个名为"NetDB2"的原型系统。在外包数据库系统中,组织将自己的数据库业务外包给外部数据库服务器运行,外包服务提供者为数据拥有者及数据库用户提供远程的数据库创建、存储、更新与查询服务。这种运行模式带来的最大挑战就是安全问题。由于数据存储在非完全可信的第三方服务器中,因此,外包数据库系统的安全机制不但要防止来自外部的恶意攻击,而且要充分考虑来自服务提供者本身的恶意操作。为了实现对外包数据库的安全控制,必须根据其结构特点,采取特定的安全机制。

本文讨论了数据库安全体系的分类,归纳了不同安全域模式下数据库系统的特点。探讨了外包数据库的基本结构,综述了外包数据库在数据加密、密文查询策略、隐私保护、数据完整性验证及基于数据库水印技术的数据库版权保护等安全机制方面的研究现状,介绍了其研究进展,并展望了外包数据库系统安全机制研究的发展方向。

1 数据库安全体系分类

根据数据库系统不同应用场合中安全域(trust domains,亦称信任域)的不同,数据库安全体系可分为4类[8],分别对应4种应用场合——单用户数据库、客户/服务器数据库、作为服务的数据库及数据发布,如图1所示。

单用户数据库系统中,数据拥有者也是数据查询的执行者,整个系统都处于数据拥有者可信的安全域,这种应用场合下的安全问题一般无碍大局。

*)2005年江苏省高校自然科学研究计划(编号:05KJD520168)、2006年东华大学科技发展基金。朱 勤 博士生,副教授,研究领域:数据库与 Web 系统安全;于守健 博士,讲师,研究领域:数据库与企业应用集成;乐嘉锦 教授,博士生导师,研究领域:数据库与数据仓库、软件工程技术;骆轶姝 博士生,讲师,研究领域:安全数据库技术。

维普资讯 http://www.cqvip.com

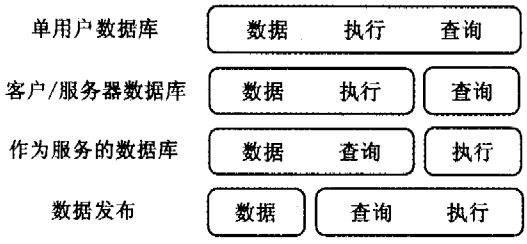

图1 基于安全域的数据库安全体系分类

客户/服务器数据库系统是网络环境下的典型应用,其安全问题讨论最多,安全机制也最为丰富。这种体系结构下,数据库服务器处于信任域,控制查询引擎,对非可信客户端的查询请求进行安全性分析并执行查询,同时进行用户鉴别、身份认证、访问控制、审计等安全控制。

数据发布模式下,数据拥有者一旦在网络上发布了自己的数据,则失去了对数据库安全的控制。用户可以下载、复制,甚至再发布数据,这样使得所有查询的提交与执行都在用户客户端进行。由于用户对数据库拥有控制权,这种模式需要着重考虑数据库的隐私保护与版权保护等问题。随着XML网络数据交换规范的逐渐形成,针对XML数据的安全数据发布技术的研究正受到越来越多的关注。

作为服务的数据库系统亦称外包数据库系统。数据拥有者位于客户端,远程更新数据;数据库用户也位于客户端,远程发布查询请求;数据存储与查询执行则依赖于提供外包数据库服务的非完全可信的第三方服务器。随着网络技术的发展与应用专业化水平的提高,这种结构模式有不断推广的趋势。本文的研究即围绕外包数据库系统的安全机制展开。

2 外包数据库系统的基本结构

外包数据库系统构建于Internet之上,充分利用了网络提供的远程访问服务。系统由3部分组成:(1)数据拥有者;(2)数据库服务提供者;(3)数据库用户。其基本结构如图2所示。

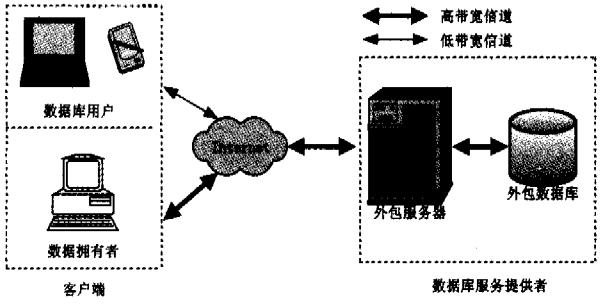

图2 外包数据库系统的基本结构

数据库服务提供者提供服务器硬件,运行标准的数据库管理系统软件,负责存储数据拥有者的数据,并提供数据库查询、更新与维护等服务。

数据库用户向数据库服务器发送数据查询请求并接受查询结果。在实际应用中,用户终端包括各种固定或移动的办公设备,如PC机、手机、无线PDA等。

数据拥有者是拥有管理权限的数据库用户,与外包数据库宽带连接,负责创建数据库、更新和维护数据库,也发布查询请求和接受查询结果。

一般地,数据拥有者与数据库服务提供者之间存在一定的信任关系。服务提供者可以复制、备份数据,以维护数据库系统的可用性和可靠性。但是,出于隐私保护的目的,外包服务提供者不应该关心数据库的内容;一些外包数据库系统采取数据库加密的方法来保证数据内容的安全,这使得数据完整性无法保证;数据作为数据拥有者多年积累的信息资源,本身具有潜在的数字权益。因此,从数据库隐私保护、数据完整性控制及数字权益保护等角度来看,对数据拥有者而言,数据库服务器是非完全可信的。这也是外包数据库系统安全控制的难点所在。

文[6]根据客户端组成形式的不同,把外包数据库系统的结构分成3类。最基本也是最简单的结构称为"统一客户端模式"(Unified Client Model)。这种模式下,外包数据库仅被单一实体使用,这个单一的客户端创建、更新和查询数据库。第二种结构称为"多查询用户模式"(Multi-Querier Model),包括两类客户端:数据拥有者和一般查询用户。其中,数据拥有者对数据库进行增、删、改等更新操作,而一般查询用户仅具有只读查询的访问权限。第三种结构最为复杂,称为"多主模式"(Multi-Owner Model),其与"多查询用户模式"的主要区别在于数据拥有者分布于多个客户端。显然,针对不同的结构模式所采取的安全机制是不一样的,其中,"多主模式"外包数据库系统对访问控制、数据完整性验证等方面的安全要求更高。

3 外包数据库系统的安全机制

外包数据库系统的安全机制包含了数据库安全的一般机制,如身份鉴别与认证、多级安全访问控制、用户跟踪与审计、数据备份与恢复等,同时,根据外包数据库系统的结构特点,还需采用特定的安全机制。

目前,对外包数据库系统安全机制的研究已经取得了一些成果,主要包括:数据库加密、密文数据库查询、隐私保护、数据完整性验证、数据库版权保护等。

3.1 数据库加密技术

由于数据库在操作系统下是以文件形式存在的,入侵者可以直接利用操作系统的漏洞窃取数据库文件,或者篡改其内容。另一方面,数据库系统的管理员可以任意访问所有数据,使得数据内容的安全无法保证。因此,对外包数据库存储的敏感数据进行加密保护,使得即使数据不幸泄露或者丢失,也难以造成泄密,有效保证了外包数据库的安全。

按照加密部件与数据库系统内核的位置关系的不同,数据库加密机制可以分为紧耦合加密与松耦合加密,分别对应于库内核加密与库外加密方式[12]。外包数据库系统中,由于数据库服务器非完全可信,加密解密都应在客户端完成。因此,库外加密方式正受到越来越多的关注。

数据库加密的粒度一般分为四种:表、属性、记录和记录属性值。总的来说,加密粒度越小,则灵活性越好,且安全性越高,但实现技术也更为复杂,对系统的运行效率影响也越大。在外包数据库系统中,为了得到较高的安全性和灵活性,采用较多的加密粒度是记录和记录属性值。

目前还没有公认的专门针对数据库加密的加密算法,因此一般根据数据库特点选择现有的加密算法来进行数据库加密。一方面,对称密钥算法的运算速度比非对称密钥算法快很多;另一方面,在公开密钥算法中,每个用户有自己的密钥对,而作为数据库加密的密钥如果因人而异,将产生异常庞大的密钥数据存储量。因此,目前数据库加密一般采用对称密

维普资讯 http://www.cqvip.com

钥的分组加密算法。

对数据库加密必然带来密钥的分配与管理问题。目前在数据库中应用比较多的是多级密钥管理体制。以加密粒度为记录属性值的三级密钥管理体制为例，整个系统的密钥由一个主密钥、每个表上的表密钥以及各个记录属性值密钥组成。表密钥被主密钥加密后以密文形式保存在数据字典中，记录属性值密钥由主密钥及记录属性值所在行、列通过某种函数自动生成，一般不需要保存。这样大大减少了密钥信息的存储量，访问效率也比集中密钥机制有了很大提高[32]。

借助于硬件执行加解密运算是近年来数据库加密技术取得的主要进展，其主要目的是提高系统的可用性。文[7]提出了一个基于硬件加密方案的外包数据库系统结构。该方案在服务器与客户端之间集成了一个用于加密解密的安全协处理器，大大降低了服务器及客户端的运行负载。文[15,16]基于智能卡(Smart Card)设计并构建客户端安全操作环境(Security Operating Environments，SOE)，执行加解密运算及访问控制，具有较高的可用性。

加密解密运算与密钥管理会显著降低数据库的访问与运行效率。外包数据库系统的保密性与可用性之间不可避免地存在冲突，需要妥善解决两者之间的矛盾。

3.2 密文数据查询策略

数据库中存储密文数据后，如何进行高效查询成为一个重要的问题。常用的密文数据库查询有两种策略：一种是不用解密而直接操作密文数据；一种是分步查询。

直接操作密文数据的应用场合包括数据库的秘密同态加密[29]、数据库的序列加密等。文[4]采用序列密码算法把表数据与随机数发生器所产生的随机数进行位异或加密。查询时，把查询词与密文异或后与随机数比较，以确定文件中是否包含该词。这种方法简单易行，几乎不增加额外存储空间，运算速度快。但是，由于序列密码算法采用异或的运算方法，密钥序列不能重复。如果对不同记录采取不同的密钥种子，则密钥管理难度太大；如果对不同记录采取相同的密钥种子，则会存在不少相同或相近的密文字段值，容易受到统计攻击和已知明文攻击[28]。

分步查询是一种更具有实用价值的密文查询策略，也是目前研究的热点。这种方法一般需要进行查询分解，先对密文数据进行范围查询，缩小解密范围，快速解密后再执行精确查询。优化的查询计划还将一次查询任务分解为客户端和服务器的多次交互，以得到最优的查询树。

Hakan Hacigumus 等提出了一种外包数据库模式下的加密数据库分步查询策略[10]，其结构模式如图 3 所示。该方法使用特定的分区函数对加密数据建立粗糙索引(Coarse Index)，存储在服务器端的加密数据库中。在查询时，把原始的基于明文数据的查询分为两步：首先运行位于服务器端的对密文数据的范围查询，利用“粗糙索引”找到包含目标记录的密文数据分区，缩小了解密范围；其次是在客户端对服务器返回的查询结果解密后的查询。文[11]进一步讨论了该模式下密文数据查询策略的优化问题。

这种查询模式具有较高的安全性和可用性。数据的完整性和一致性由服务器负责，服务器的所有操作都是针对加密数据；网络传输密文数据，数据的加密和解密都在客户端完成，数据的安全性由客户端控制。查询策略的核心在于尽量提高对密文数据库查询的准确率，缩小返回客户端的密文数据的范围。

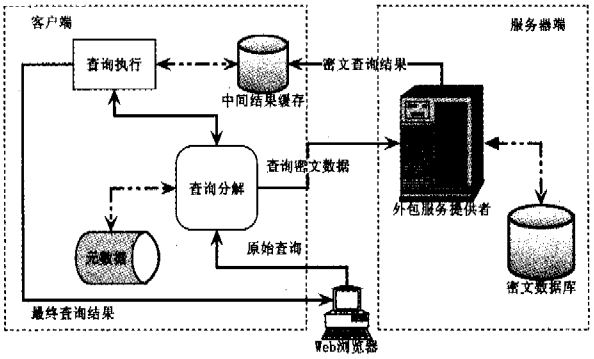

图 3 外包数据库的密文数据查询模式

目前密文数据查询技术面临的主要困难是：算法执行效率不高；大多不能支持模糊匹配查询；在处理多表查询时，会产生大量的伪连接，解密工作量大大增加；在处理多条件查询时，由于大量中间结果的产生，导致系统运行效率下降。

总的来说，密文数据库的查询是一项复杂的工作，目前还处于研究阶段。高效、安全的密文数据库查询策略是外包数据库系统安全技术研究的热点之一。

3.3 数据库隐私保护

隐私被认为是个体来控制自己的信息的权力。越来越多的因故意或疏忽造成的信息泄露的例子，使人们对数据库中的隐私保护问题日益重视。

外包数据库的隐私保护包括两个方面：一方面是对以明文存储的隐私数据内容的保护；另一方面是对用户查询行为及查询结果的保护，即保密信息检索(Private Information Retrieval，简称 PIR)。

3.3.1 基于推理控制的隐私内容保护

数据库用户往往能够根据低密级的数据和模式的完整性约束推导出高密级的数据，造成未经授权的信息泄漏，这种推理的路径称为推理通道(Inference Channel)。

常见的推理通道主要包括：利用多次查询的结果之间的逻辑联系进行推理；利用不同安全级别数据之间的函数依赖进行推理分析；利用数据完整性约束进行推理。

目前常用的推理控制方法可以分为 4 种：语义数据模型方法、形式化方法、多实例方法和查询限制方法。

Hinke 在 ASD Views 工程中构建了一个语义关系图来表达数据库中可能的推理[33]。数据项被表示为节点，它们之间的关系由连接节点的边表示。如果两个节点之间有两条路径，并且从一条路径可以看到所有的边，而从另外一条路径却不能，那么就可能存在一条推理通道。相应的解决方法是提升边的级别，直到所有的推理通道被关闭。

文[25]给出了消除函数依赖和多值依赖推理的形式化算法。对于函数依赖推理，采用提高属性的安全级的算法来消除推理通道；对于多值依赖推理，把关系实例中的相应元组的安全级升高，经过元组的安全级调整后，新的关系实例不再存在多值依赖推理。

多实例是指数据库中允许存在关键字相同但安全级别不同的元组，即把安全级别作为主关键字的一部分，从而解决了利用主关键字的完整性进行推理的问题[33]。多实例方法的缺点是使数据库失去了实体完整性，同时增加了数据库中数据关系的复杂性。

文[23]提出了查询限制的方法。当系统接受到用户提交的查询的时候，首先判断该查询是否可以导致敏感信息的推

维普资讯 http://www.cqvip.com

理，如果可以，则对查询进行转换；文[3]假定数据库由一个全局关系组成，全局关系可以通过所有关系的笛卡尔积得到。从查询、谓词和模式在表示数据库元组等价的角度出发，通过检查 SQL 语言的 Where 子句，可以发现大部分推理通道。

以上几类方法并不是相互孤立，而是可以综合运用，从整体上达到消除更多推理通道和保护隐私的目的。文[13]建立了信息泄漏的表语义与查询语义模型，然后通过修改 SQL 语言查询条件的方法来进行查询预处理，实现了数据元素粒度的推理控制。

一般来说，推理控制可以防止由推理得到未经授权的存取路径，也可能会限制合法用户的正常访问，在实际系统实现时需要权衡两方面的合理需求。

3.3.2 保密信息检索

保密信息检索的需求源自外包数据库服务器的不可信。用户向服务器提交查询请求，同时不希望对服务器公开他获取的查询结果。

保密信息检索在一些特定的数据库应用场合有重要意义。如军事行动部门在执行一个特别行动任务时，会事先从 IT 部门的数据库服务器查询目的地的地图等相关信息。显然，这样的查询不希望被服务器端知晓。再如股票交易中，用户对股票数据库的查询往往被认为是敏感的，需要加以保密。

目前，保密信息检索的解决方案主要有两类：多服务器方式和单服务器方式。

多服务器方式将数据库复制在相互之间不可通信的多个服务器上[2]。用户检索数据时，经过算法处理生成的多个查询请求被发往多个服务器，任何一个服务器都不能唯一确定用户的查询结果，而客户端将返回的多个查询结果组织为用户所需要的数据。文[24,1]进一步讨论了减少运算复杂度和通信负载的多服务器方式。

单服务器方式省去了多个数据库复制的麻烦，其基本思路是提交加密的查询请求[14]。服务器运行特定的算法处理这些加密查询，而查询结果只有客户端才能够解密。文[PLH04]提出了一种对树型结构数据进行保密查询的解决方案，其主要思路是冗余访问和节点交换，使得服务器对被查询数据的遍历路径不可知，从而实现了保密信息检索。

保密信息检索技术面临的主要问题是计算复杂度和通信开销都比较大，需要进一步进行算法的优化，在保证安全性的前提下提高可用性。同时，现有的协议与算法一般只考虑了简单的选择查询，对于多条件组合查询、聚集操作等复杂操作的保密查询还有待进一步的研究。

3.4 数据完整性验证

外包数据库系统的数据完整性，要求数据库内容及其在网络中的传输具有正确性、一致性与有效性，以确保接收到的数据库内容是真实有效的，并且在传输过程中没有被攻击者插入、篡改、伪造、重排等。在电子商务、电子政务中，数据完整性是最基本、最重要的安全要求。

如果仅仅考虑来自外部的攻击，即确保客户端与服务器之间的通信不受破坏，那么只要采用常规的安全通信协议，如 IPSec(安全 IP 协议)、SSL(安全套接字层协议)等，即可满足需求。但是对外包数据库系统来说，这并不能保证来自服务器的数据真实有效。因为，在外包数据库系统中，服务器是非完全可信的，它并不关心数据的完整性，来自服务器内部的恶意操作可能会对数据库添加虚假记录，或者篡改已有记录。因此，外包数据库的数据完整性要求的特点是，既要考虑外部攻击，又要考虑服务器自身的安全性。

数据库的数据完整性粒度可以分为 4 种：表完整性、属性完整性、记录完整性及记录属性值完整性。考虑到外包数据库中频繁的远程查询操作，每次验证表完整性所带来的繁重运算量，以及维护记录属性值完整性所需的巨大存储开销，目前外包数据库中维护数据完整性一般采用记录粒度，即对每条记录的存储、更新与查询进行完整性验证。

实现数据完整性的主要措施就是增加攻击者所不能控制的冗余信息。外包数据库中，根据系统结构模式不同，所采取的完整性验证机制也不相同。对于“统一客户端模式”，采用加密 Hash 函数，如消息验证码(Message Authentication Codes，MAC)等，对数据库记录进行数字签名[21]。由于数据拥有者与查询用户为同一客户端，使用同一密钥就可以安全地进行完整性验证。而对于“多查询用户模式”及“多主模式”，数据的更新与查询来自多个用户，显然不能共享密钥，这无疑增加了维护数据完整性维护与验证的复杂度。

文[6]提出了一种基于公钥机制的压缩 RSA 算法(Condensed-RSA)的批量数字签名方案，适用于“多查询用户模式”的数据完整性控制。利用 RSA 算法的倍增同态性，一方面，可以取出与查询结果集相对应的多个数字签名，组合为一个“压缩签名”连同查询结果一起发送给查询者；另一方面，查询者可以根据其访问权限，方便地从组合签名中鉴别出单个的签名者。这样较好地满足了多用户查询条件下的数据完整性要求，同时，其运算、存储与通信负载也在可接受的范围之内。

文[6]同时提出了采用群签名(Group Signature)算法以及 Merkle Hash 树(MHT)批量签名的解决方案，满足“多主模式”的数据完整性要求。文[17]给出了一个名为“数字签名组合与链接”(Digital Signature Aggregation and Chaining，DSAC)的解决方案，使“多主模式”下的外包数据库完整性控制的可用性向前迈进了一步。

从现有的研究成果来看，外包数据库“多主模式”的数据完整性验证涉及较高的计算复杂度，离实际应用还有相当一段距离。同时，现有的解决方案一般仅能支持基本的数据库查询语句，而对求和等聚集操作所查询数据的完整性验证问题还有待进一步研究。

3.5 外包数据库版权保护

在外包数据库服务模式中，数据经过若干年的积累，往往蕴含有巨大的社会价值与经济价值，成为宝贵的数据资源。而由于数据库服务器由第三方提供，数据库的物理文件可以轻易地被拷贝而造成资源盗用，因此，数据拥有者对数据库实施版权保护的需求日益迫切。

2000 年，S Khanna 等提出利用数字水印实现对数据库安全控制的新思路[22]，使数据库水印技术引起研究者关注。数据库水印是指用信号处理的方法在数据库中嵌入不易察觉且难以去除的标记，在不破坏数据库内容和可用性的前提下，达到证明数据拥有者版权的目的。由于数据库的特殊性，在其中不易找到能插入水印标记的可辨认冗余空间，因而研究具有一定难度，成熟的水印算法还很少。

美国 IBM Almaden 研究中心的 Rakesh Agrawal 等在这方面作了开创性的研究，他们于 2002 年首次进行了向关系数据库嵌入比特位模式的实验[18]。该实验利用数据库关系中数值型元组存在的冗余空间，对其最低有效位进行位操作，实现一位水印版权信息的嵌入。

维普资讯 http://www.cqvip.com

美国 Purdue 大学的 Radu Sion 等提出了基于元组排序和划分集合实现水印嵌入的算法[19]。首先根据元组的加密键值哈希对其进行秘密排序，然后基于“均方差”特性构造子集，取连续序列数据作为嵌入水印的基本单位，通过调整关键属性数据改变连续序列数据的分布特征来表示 1 和 0。基于该算法开发的用于数据库版权保护的水印关系数据库程序包“WMDB”已经投入应用[27]。

上述两种数据库水印算法各有其限制特点。第一类方法采用基本的 LSB 嵌入算法，易于实现，但水印信号的抗攻击能力较弱，而且难以嵌入有实际意义的水印信息。第二类方法具有较好的鲁棒性，但如果数据库中不同字段的取值范围相差比较大，将导致计算获得的值只能对部分数据项适用，限制了水印嵌入的容量。另外，该算法对于频繁更新的数据库的水印嵌入开销巨大。

目前投入实际运行的水印数据库系统还不多见。Elisa Bertino 等提出了一个综合隐私保护和版权保护功能的外包数据库系统的框架[5]。该系统基于领域等级树(Domain Hierarchy Tree，DHT)和 k-匿名(k-anonymity)算法实现隐私保护，应用数字水印技术实现数据库的版权保护。不足之处在于：其保护对象局限于在逻辑上具有层次关系的隐私数据库，同时对于数据更新后的水印运算开销也很大。

利用数字水印实现对外包数据库的版权保护，具有较高的研究和应用价值。目前，数据库水印技术在水印宿主数据类型扩展[20]、XML 数据水印[26]、数据库脆弱水印[30]、数据库数字指纹[31]等研究方向已取得一定的进展。随着研究的深入，数据库水印将获得更好的透明性、鲁棒性与可用性，其在外包数据库安全领域的应用空间必将得到进一步拓展。

4 外包数据库安全研究的发展方向

未来几年，外包数据库系统安全技术的研究可能会在以下几个方面得到发展：

(1)形式化安全策略模型的完善。形式化方法，包括建模、证明、分析与应用逻辑表达等，是数据库安全理论与应用的重要基础，在实际系统应用中的修正与扩展是抽象模型获得实用性的必要环节。目前还不能严格地形式化证明一个系统的安全强度。只有将完整性、保密性、真实性等逻辑有机集成到形式化的安全模型中，外包数据库系统的安全性才能得到可靠的理论保证。

(2)数据库系统安全机制与操作系统和网络的安全机制的进一步整合。数据库系统是运行于操作系统与网络环境之中并依赖其提供的安全功能和机制的，因此外包数据库系统与安全操作系统以及安全网络环境相结合，才能达到整体的安全性。操作系统安全和网络安全的一些新的研究成果，如双重身份认证、单点登录、入侵检测技术等，应有效地集成到外包数据库系统的安全机制中来。

(3)数据库水印等新技术在外包数据库安全控制中的应用拓展。数据库水印是一种从非加密的角度实现对数据库安全控制的新思路，目前的应用研究还只限于数据库版权保护。结合传统的密码学机制，将数据库水印应用于数据库的访问控制、入侵检测、隐私保护等领域，将使外包数据库系统的可用性大大提高。

结束语 以上探讨了外包数据库的多种安全机制。必须说明的是，这些安全技术不是相互独立而是相互支持的，而且一般要结合常规的数据库安全技术，以提高系统的整体安全性。同时，数据库系统的安全性和可用性始终是一对矛盾，在保证安全性的前提下，需要进一步提高系统的可用性，以构造出一个既具方便使用、又能最大限度地提供安全保障的外包数据库系统。

总的来说，外包数据库系统的安全机制正处于研究与发展当中，尚有许多重要的技术问题需要解决。随着电子商务、电子政务中数据库外包服务需求的不断增多，以及对系统安全性要求的不断提高，对它的研究不仅具有重要的理论价值，而且对国家安全和国民经济的健康发展具有重要意义。

参 考 文 献

1 Beimel A, Ishai Y. Information-Theoretic Private Information Retrieval: A Unified Construction: [ECCC Report TR01-015]. Feb. 2001

2 Chor B, Goldreich O, Kushilevitz E, et. al. Private Information Retrieval. In: Proc. of 36th IEEE Conference on FOCS, 1995

3 Marks D G. Inference in MLS Database Systems [J]. IEEE Transaction on Knowledhe and Data Engineering, 1996, 8(1): 46～55

4 Song Xiaodong D, Wanger D, Perring A. Practical Techniques for Searches on Encrypted Data. In: The IEEE Symp on Security and Privacy, Oakland, California, 2000

5 Bertino E, Ooi Beng Chin, Yang Yanjiang, et al. Privacy and Ownership Preserving of Outsourced Medical Data. ICDE, 2005. 521～532

6 Mykletun E, Narasimha M, Tsudik G. Authentication and Integrity in Outsourced Databases. In: Proceedings of NDSS 2004, San Diego, California, USA

7 Mykletun E, Tsudik G. Incorporating a Secure Coprocessor in the Database-as-a-Service Model. In: IWIA2005 for Future Generation High Performance Processors and Systems, Oahu, Hawaii, USA, Jan. 2005

8 Miklau G, Suciu D. Controlling Access to Published Data Using Cryptography. In: VLDB, 2003

9 Hacigumus H, Iyer B, Mehrotra S. Providing Database as a Service. In: Proc. of ICDE, 2002

10 Hacigumus H, Iyer B, Mehrotra S. Executing SQL over Encrypted Data in Database Service Provider Model. In: Proc. of ACM SIGMOD, 2002

11 Hacigumus H, Iyer B, Mehrotra S. Query Optimization in Encrypted Database System. In: Proc. of DASFAA, 2005, LNCS 3453. 43～45

12 He J, Wang M, Cryptography and Relational Database Management Systems. In: Database Engineering and Application Symposium, 2001. 273～284

13 LeFevrey K, Agrawal R, Ercegovac V, et al. Limiting Disclosure in Hippocratic Databases [C]. In: Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004

14 Kushilevitz, Ostrovsky R. Replication is NOT Needed: Single-database Computationally Private Information etrieval. In: Proc. of 38th FOCS, 1997

15 Bouganim L, Pucheral P. Chip-Secured Data Access: Confidential Data on Untrusted Servers. In: Proceedings of the 28th Very Large Data Bases Conference, Hong Kong, 2002. 131～142

16 Bouganim L, Cremarenco C, Ngoc F D, et al. Safe Data Sharing and Data Dissemination on Smart Devicesrequires. In: SIGMOD 2005, Baltimore, Maryland, USA

17 Narasimha M, Tsudik G. Authentication of Outsourced Databases Using Signature Aggregation and Chaining. In: International Conference on Database Systems for Advanced Applications (DASFAA 2006), Apr. 2006

18 Agrawal R, Kiernan J. Watermarking Relational Databases [A]. In: Proceeding of the 28th VLDB Conference, 2002

19 Sion R, Atallah M, Prabhakar S. Rights protection for relational data [A]. In: 2003 ACM SIGMOD International Conference, 2003. 98～109

20 Sion R, Atallah M, Prabhakar S. Ownership Proofs for Categorical Data. IEEE Journal of Transactions on Knowledge and Data Engineering IEEE TKDE, 2005, 17(7)

(下转第 195 页)

维普资讯 http://www.cqvip.com

的任意形状，并要求获得整个数据集的全局信息，这在内存有限情况下对动态数据集进行聚类是难以适用的。如何在有限内存情况下对动态数据集进行增量式聚类是我们下一步研究工作的方向。

参 考 文 献

1 Ester M, Kriegel H -P, Sander J, et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: Proc. 2nd Int Conf on Knowledge Discovery and Data Mining, Portland, OR, 1996. 226～231

2 Han Jiawei, Kamber M, Fan Ming, et al. 数据挖掘：概念与技术. 北京：机械工业出版社，2001

3 Ankerst M, Breunig M, Kriegel H-P, et al. OPTICS: Ordering Points To Identify the Clustering Structure. In: Proc. ACM SIGMOD'99, Int Conf. on Management of Data, Philadelphia, PA, 1999

4 Breunig M M, Kriegel H-P, Ng R T, et al. LOF: identifying density-based local outliers. In: Proc. ACM SIGMOD 2000 Int Conf on Management of Data, Dalles, TX, 2000

5 Tang Jian, Chen Zhixiang, Ada Wai-chee Fu, et al. A Robust Outlier Detection Scheme for Large Data Sets. In: http://www.cs.panam.edu/～chen/papers.html

6 Zhou Yong-Feng, Liu Qing-Bao, Deng Su, et al. An Incremental Outlier Factor Based Clustering Algorithm. In: the First International Conference on Machine Learning and Cybernetics, Nov2002, CHINA

7 Jin Wen, Tung A K H, Han Jiawei. Mining Top-n Local Outliers in Large Databases. In: Proc. ACM KDD 2001, San Francisco, California USA

（上接第 156 页）

21 Sion R. Query Execution Assurance for Outsourced Databases. In: Proceedings of the 31st VLDB Conference, Trondheim, Norway, 2005

22 Khanna S, Zane F. Watermarking maps: hiding information in structured data. SODA, 2000. 596～605

23 Thruaisingham B. Recursion Theoretic Properties of the Inference Problem in Database Security [R]. In: MTP291. MITRE Corp, Bedford, Mass, 1990. 21～33

24 Itoh T. On Lower Bounds for the Communication Complexity of Private Information Retrieval. IEICE Transactions, 2001, E84- A (1)

25 Su T, Ozsoyoglu G. Controlling FD and MVD Inferences in Multilevel Relational Database System [J]. IEEE Transactions on Knowledge and Data Engineering, 1991, 3(4): 474～485

26 Ng W, Lau Ho-Lam. Effective Approaches for Watermarking XML Data. In: Proc. of DASFAA, LNCS, 2005, 3453: 68～80

27 WMDB System Architecture. http://www.cs.stonybrook.edu/～sion/projects/wmdb, 2004

28 王正飞，王曼，汪卫，等. 数据库中加密字符数据的存储与查询 [J]. 计算机研究与发展，2004, 41(Suppl. 10): 66～71

29 王晓峰，王尚平. 秘密同态技术在数据库安全中的应用 [J]. 计算机工程与应用，2003, 14: 194～196

30 Li Yingjiu, Guo Huiping, Jajodia S. Tamper Detection and Localization for Categorical Data Using Fragile Watermarks. In: Proceeding of the DRM'04, October 25, 2004, Washington, DC, USA

31 Li Yingjiu, Swarup V. Fingerprinting Relational Databases - Schemes and Specialties. IEEE Transaction on Dependable and Secure Computing, 2005, 2(1): 34～45

32 张敏，徐震，冯登国. 数据库安全 [M]. 北京：科学出版社，2005. 163～169

33 朱虹，史凌云，张勇. 多级安全数据库系统推理问题研究 [J]. 计算机工程与应用，2004, 13: 179～181

（上接第 165 页）

人名的组字相对来说比较稳定，地名和机构名也有比较好的后缀特征可以利用。如果把概念词抽取中把专名识别独立地作一个模块，利用其中独有的一些规律，对专名进行推测和验证，系统的准确率会更高。

（2）概念词内部语义关系的获取：分析复合概念词与构成它的基本概念词之间的语义关系，从而为领域 Ontology 的辅助构建、概念词辅助定义、概念词的辅助翻译提供支持。

（3）概念词关系的获取：概念之间存在着上下位关系、部分关系、同指关系、同义关系、反义关系等关系；利用概念词内部的语义信息和语料中的分布特点，可以进一步获取概念词间的关系。

参 考 文 献

1 Bourigault D. Surface Gramatical Analysis for the Extraction of Terminological Noun Phrases. In: Proceedings of COLING 92. 977～981

2 Frantzi T K. Incorporating Context Information for the Extraction of Terms. In: Preceedings of ACLEACL' 97

3 Wu shih-hung, Hsu wen -Lian. A semi-automatic domain ontology acquisition tool from Chinese Corpus [C]. In: Proc. of the 19th International Conference on Computational Linguistics (COLING)2002, Taipei, Taiwan, 2002. 1313～1317

4 Dunning T. Accurate Methods for the Statistical of Surprise and Coincidence. Association for Computational Linguistics, 1993, 19 (1): 61～76

5 Pantel P, Lin Dekang. A Statistical Corpus-based Term Extractor. In: Canadian Conference on AI, 2001. 36～46

6 Enguehard C, Pantera L. Automatic Natural Acquistion of a Terminology. Journal of Quantitative Linguistics, 1994, 2(1): 27～32

7 Luo S F, Sun M S. Two-Character Chinese Word Extraction Based on Hybrid of Internal and Contextual Measures. In: Proceeding of ACL2003, Sapporo, Japan, 2003

8 Riloff E. Automatically constructing a dictionary for information extraction tasks. In: Proceedings of the Eleventh National Conference on Artificial Intelligence, 1993. 811～816

9 郑家恒，杜永萍，宋礼鹏. 农业病虫害词汇获取方法初探. 见：孙茂松，陈群秀. 语言计算与基于内容的文本处理. 北京：清华大学出版社，2003. 61～66

10 张春霞. 领域文本知识获取方法研究及其在考古领域中的应用：[博士论文]. 北京：中科院计算所，2005

11 罗贝，吴洁，曹存根，等. 从文本中获取植物知识方法的研究. 计算机科学，2005, 32(10): 6～13

12 刘磊，曹存根. 一种基于"是一个"模式的下位概念获取方法. 计算机科学(已录用). 2006

13 http://icl.pku.edu.cn/icl%5Fres/segtag98/