数据库服务中基于密码学的访问控制

◎ 田秀霞 孙建国 周傲英 顾春华 著

清華大學出版社

数据库服务中基于密码学的访问控制

◎ 田秀霞 孙建国 周傲英 顾春华 著

清華大學出版社
北京

内 容 简 介

本书是针对安全研究领域的学术专著，内容包括基于属性加密的访问控制增强、基于谓词加密的访问控制增强、基于代理加密的访问控制增强、基于密码提交协议的访问控制增强，以及基于密文策略属性集合加密的访问控制增强。

本书适用于信息安全领域或数据库安全领域的研究者。此书在多项基金资助下撰写完成：国家自然科学基金(No. 61202020)、CCF-腾讯犀牛鸟创意基金(No. CCF-TencentlAGR20150109)、上海市科委地方能力建设基金(No. 15110500700)。

图书在版编目(CIP)数据

数据库服务中基于密码学的访问控制/田秀霞等著. —北京：清华大学出版社，2017
ISBN 978-7-302-43916-5

Ⅰ. ①数… Ⅱ. ①田… Ⅲ. ①数据库－密码－访问控制 Ⅳ. ①TP309.7

中国版本图书馆 CIP 数据核字(2016)第 111160 号

责任编辑：付弘宇　薛　阳
封面设计：刘　键
责任校对：梁　毅
责任印制：刘海龙

出版发行：清华大学出版社
网　址：http://www.tup.com.cn，http://www.wqbook.com
地　址：北京清华大学学研大厦 A 座　　邮　编：100084
社总机：010-62770175　　邮　购：010-62786544
投稿与读者服务：010-62776969，c-service@tup.tsinghua.edu.cn
质　量　反　馈：010-62772015，zhiliang@tup.tsinghua.edu.cn
课　件　下　载：http://www.tup.com.cn，010-62795954

印 装 者：北京泽宇印刷有限公司
经　　销：全国新华书店
开　　本：170mm×230mm　　**印　张**：7.75　　**字　数**：122 千字
版　　次：2017 年 2 月第 1 版　　**印　次**：2017 年 2 月第1次印刷
印　　数：1～1000
定　　价：35.00 元

产品编号：066577-01

前言
FOREWORD

随着越来越多的数据以电子的形式被收集、存储，许多企业开始难以承受海量数据管理和维护带来的人力、财力、物力等巨大开销，转而希望将自己的海量数据委托给一个既能提供基本的、可靠的硬件基础设施，又能提供专业数据管理的第三方服务提供者存储、管理和维护。数据库服务作为一种新的基于云计算平台的网络数据管理模式满足了企业的这种需求，并可以提供像本地数据库一样的数据管理服务。然而，越来越多的数据涉及敏感信息，如医疗记录、交易信息、证券信息、财务信息等，此外，企业间的竞争以及数据库隐私数据窃取或泄漏促使企业必须选择具有安全和隐私保护能力的网络数据管理技术。

有关数据库服务中安全技术的研究已有多年，但大多在数据的机密性、数据的完整性、数据的完备性、查询隐私保护等方面对数据库服务进行研究，而对隐私保护机制，如提高密文数据库可用性的访问控制增强、保护用户和策略隐私的访问控制增强等方面研究较少。

本书主要从以下几个方面论述基于密码技术的访问控制增强，实现数据库服务模式下隐私增强的访问控制机制。

(1) 数据库加密：

① 采用不同的加密机制加密数据库数据，如对称加密、非对称加密、代理加密、基于属性的加密、谓词加密等；

② 设置不同的加密粒度，如文件级、表级、元组级、属性级；

③ 设置不同的加密层次，如存储层加密、数据库层加密、应用层加密。

(2) 基于属性加密的访问控制增强：访问控制策略根据合法用户属性的特性描述，而不是消费者用户的真实身份等，再有就是能够访问数据特定区块的授权消费者用户列表也是不能事先知道的。特别是基于属性的访问授权，提供了更好的表达性和可控性。

(3) 基于谓词加密的访问控制增强：在第三方提供外包管理服务的数据库服务场景中，数据库所有者可能想定义一个策略来决定谁可以恢复秘密数据。如分类的数据可能和特定的关键词关联，这些数据可以被允许阅读所有类信息的用户访问，也可以被允许阅读和特定关键词关联的用户访问，再如，医疗外包数据能被具有不同访问许可的医生、个人或疾病研究机构访问。在第三方服务提供者仅提供检索转发服务的应用中，邮件转发服务器只需要检测加密的邮件是否满足用户查询的邮件关键词，而不需要知道加密的邮件内容。谓词加密作为一种新的密码学机制提供了密文数据上细粒度的访问控制。

(4) 基于密码提交协议的访问控制增强：实现了在数据库服务提供者端增强数据库拥有者的访问控制策略，同时也保证了用户身份和委托访问控制策略的隐私。该机制利用密码学中被证明是无条件隐藏的 Pedersen 协议来保证用户身份属性的隐私；分别用根据访问控制策略的选择加密和根据委托访问控制策略的选择加密来增强数据库拥有者端和数据库服务提供者端的选择访问授权。

(5) 基于密文策略属性集合加密的访问控制增强：采用基于密文策略属性集合加密和数据库服务提供者再加密相融合的方法实现。提出的机制实现了多重隐私保证下的访问控制增强：委托数据中的数据隐私、委托授权表中的策略隐私和密钥分布过程中的密钥隐私。

目录
CONTENTS

第1章

绪　论

1.1　数据库服务应用背景

随着网络技术和通信技术的飞速发展，越来越多的应用如电子银行、社交网络、电子政务、医疗服务等发展成网络应用，通过这些应用收集的消费者信息、交易信息、医疗信息等海量数据以电子化的方式被存储和管理。这使得很多企业不得不高薪聘请专业的数据库管理人员进行数据库的安全管理、灾难恢复、软件更新等非企业核心业务工作。为了将自身从非核心的数据库管理业务中解脱出来，进而专心于创造更大的核心业务价值，越来越多的企业开始寻求一种专业的数据库管理服务，并希望其能够代表企业利益提供如同本地管理一样的数据维护、软件更新、灾难恢复等数据管理服务。

本章参考文献[1]提出了数据库服务的概念，该技术一经提出就受到来自学术界和工业界的广泛关注[22]，如 Amazon 关系数据库服务（Amazon Relational Database Service，Amazon RDS）、基于 Microsoft Azure 的 SQL 数据库（Microsoft Azure SQL Database，MS SQL）、Google 的 Cloud SQL，Salesforce 的 Database. com、Baidu 云数据库、Aliyun 关系型数据库等。工业界基于云数据库管理服务的实际部署进一步验证了企业的需求

趋向。

1.2 数据库服务基本概念

由于越来越多的数据信息涉及用户个人隐私或相关隐私，如在线银行中的用户信息(如信用卡号和密码)，在线医疗中的敏感病症(如传染病或癌症)，在线社交网络中的朋友圈(如富有朋友圈)等。而一系列隐私数据泄漏事故[23]，导致大量网民受到隐私泄漏的威胁，如已经成为规模经济的信用卡黑色地下交易链；2011 年 12 月 CSDN、世纪佳缘等多家网站的用户数据库被曝光在网络上，部分密码以明文方式显示；2013 年 10 月，如家、七天等连锁酒店被网曝有多达 2000 万条客户开房信息遭泄漏；2013 年 11 月，圆通速递近百万条快递单个人信息在网络上被公开出售，网上甚至还出现了专门交易快递单号的网站；2014 年 12 月，12306 火车订票 13 万用户信息泄漏事件；2015 年 5 月，携程物理数据库被删除等。因此，这里参考在综述文献[2]和学位论文[3]中提出的数据库服务概念，并结合工业界[22]对数据库服务的功能需求以及不同实体的隐私保护需求，补充给出如下数据库服务概念(本文之后出现的相关概念和符号都以该定义为参考基准)。

【定义 1：数据库服务】 数据库服务(Database as a Service，DaaS)，也称作数据库外包(Database Outsourcing)，就是指企业(数据拥有者)将自身的数据库创建、访问、维护、升级、管理、安全设置等任务委托给专门的可以提供这些功能的第三方(数据库服务提供者)管理。采用 DaaS 实现的管理优势：一方面可以减轻企业购买昂贵的软件、硬件、处理软件升级、雇用数据库管理和专业维护人员耗费的负担，另一方面，企业也可以将有限的资源集中在自身具有核心竞争力的业务上。同时，提供 DaaS 的专业企业也可以通过取得大量该业务的订单，对不同的企业提供类似的服务来减小开支，取得规模经济，获得利润。采用 DaaS 需要增强隐私保护需求：一方面提供数据机密性、完整性、访问控制授权等基本安全需求，另一方面结合数据的敏感性和用户隐私需求提供特定的数据隐私(数据对 DSP 是不可见的)、请求用户(数据消费者)的查询保护(如完备性、查询隐私等)、访问控

制隐私增强等。

1.3 数据库服务框架

我们在文献[2]中用面向服务(Service-Oriented Architecture,SOA)的观点来刻画了数据库服务的架构,如图 1.1 所示。该架构主要包括三个角色(数据拥有者、数据库服务提供者和数据请求者)和三类数据(数据源、查询与结果、密钥)。下面对这三个角色分别给予介绍。

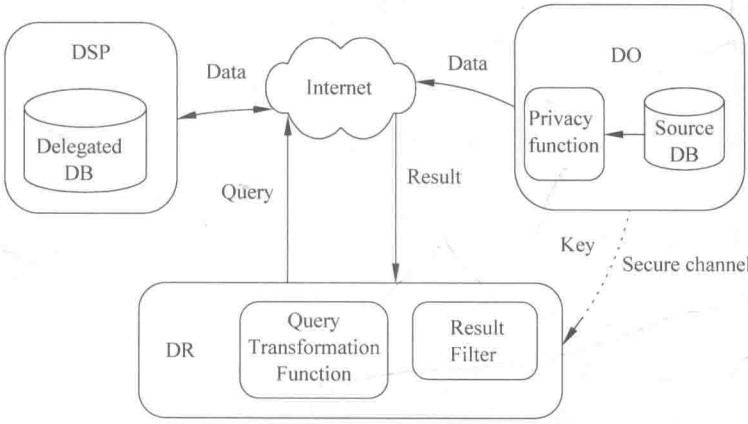

图 1.1 DaaS 系统架构

(1) 数据库服务提供者(DaaS Provider,DSP)。DSP 是指一个专业的提供 DaaS 的企业(如 Amazon,Microsoft 等),维护客户的数据库(图 1.1 中委托管理的数据库(Delegated DB)),并能够像本地数据库一样正确地进行数据库的复制、备份等数据管理任务。但是 DSP 不一定能够保证委托数据的机密性,并且其本身也可能是数据库的攻击者。所以为了防止 DSP 对委托的数据库中敏感数据的未授权访问,他从 DO 处接收的数据是经过保护数据隐私方式(如图 1.1 中 DO 处的隐私数据处理模块(Privacy function))处理过的数据。DSP 可以根据 DO 提供的辅助信息(索引信息)在不泄漏数据隐私的情况下有效地响应数据请求者的查询,而且不能解密密文数据并查看数据请求者真实的查询结果。

(2) 数据拥有者(Data Owners,DO)。DO 是实际上拥有自身用户数

据(图 1.1 中数据源(Source DB))的企业或个人(如使用第三方存储平台的个人用户),收集用户数据,并将用户数据以保护数据隐私的方式委托给 DSP。为了加快对委托的密文数据库的查询效率,DO 需要提供一些辅助手段,如针对某些字段建立保护隐私的索引或采用保护隐私的访问控制授权等,提高委托的密文数据库的可用性。

(3) 数据请求者(Data Requestor,DR)。DR 是指可以将用户的查询转换成数据库服务器可识别的查询(如通过图 1.1 中 DR 处的查询转换(Query Transformation Function)实现)、将数据库服务器返回的保护隐私的查询结果经过处理(如通过图 1.1 中 DR 处的结果过滤(Result Filter)实现),方便用户进行查询后处理的前端。DR 具有一定的计算和存储能力,如 DR 可以是计算机,也可以是移动电话或无线 PDA 等。

图 1.1 中涉及的三类数据传送操作是 DO 和 DSP 之间的数据传送、DR 和 DSP 之间的查询和结果返回以及 DO 和 DR 之间的密钥分发和验证结构传送,具体如下。

(1) DO 和 DSP 之间的数据传送。DO 和 DSP 之间的数据传送是指在传统数据库的基础上增加的 DO 和 DSP 之间的交互。在 DO 和 DSP 之间传送数据时,数据必须以某种保护数据隐私的方式(如加密数据)传送,因为企业传送的数据可能涉及财务、用户身份、客户资料等隐私信息。

(2) DR 和 DSP 之间的查询和结果返回。DR 和 DSP 之间的查询和结果返回是指 DR 可以向 DSP 提交查询,提交的查询和客户/服务器模式类似,不同的是 DR 的查询需要经过查询转换,转换成 DSP 可以识别的相关属性的隐私保护形式。DSP 接受查询并在密文数据库上执行该查询,然后在不泄漏数据隐私的情况下返回查询结果。

(3) DO 和 DR 之间的密钥分发和验证结构传送。DO 和 DR 之间的密钥分发和验证结构(一般采用参考文献[40]中的方法构造验证对象)传送,主要是为了使得 DR 能够验证 DSP 正确并且完备地返回了 DO 希望返回的数据,DO 将验证的密钥或验证结构以一定的安全方式(如图 1.1 中虚线表示的安全信道(Secure channel))传送给 DR。

研究学者 Hacigumus[1],Mykletun[27] 也对 DaaS 架构进行了一些探索。Mykletun[27] 提出了三种模式:统一客户模式、多查询者模式和多数据

拥有者模式。统一客户模式是指每一个委托的数据库仅有一个客户(上述数据拥有者和数据请求者归一)使用,该客户创建、维护和查询数据等。多查询者模式是指与图 1.1 匹配的模式,有两种类型的客户(分别是图 1.1 中的 DO 和 DR),DO 添加、删除和修改数据库记录,而多个数据请求者可以查询数据库。多数据拥有者模式是指可以有多个拥有不同安全原则的数据库拥有者创建数据库,并将数据库委托给数据库服务提供者管理和维护。实际上,后面两个模型的共同点是都可以存在多个数据请求者。

1.4 数据库服务中安全机制新挑战

数据库服务虽然可以为客户提供必要的硬件、软件维护等,但是由于越来越多的数据信息涉及个人隐私,如一个人是否患有不希望公开的传染病或癌症,而且企业间的竞争以及数据库隐私数据窃取或泄漏促使企业选择具有安全和隐私保护能力的数据库管理技术,因此,目前对数据库服务安全技术的研究主要集中于以下几个方面:数据的机密性、数据的完整性、数据的完备性、查询隐私保护和访问控制策略。每个方面涉及的研究内容以及不同于客户/服务器模式的安全机制新挑战总结描述如下。

(1) 数据的机密性。数据库的内容往往涉及企业的隐秘信息,因此数据库服务需要有完善的数据安全机制来保证数据库的内容不会泄漏(数据的机密性)。数据的机密性,由图 1.1 中 DO 处的隐私数据处理模块(如加密、数据分布等)实现,是指 DO 在将其数据委托给 DSP 之前需要对被委托的数据进行隐私保护处理,经过处理的数据可以保证数据库的内容在没有授权的情况下不能被访问,包括 DSP 在内,或者即使可以访问也因为不知道加密数据的密钥而不能推导出真实的委托数据。DaaS 提供的数据的机密性主要包括两层含义:一是保护数据不被未授权的 DR 访问;二是保护数据不被不可信的 DSP 访问。只有这两种情况下的机密性都被保证的情况下才可以保证企业机密的信息不会泄漏。

(2) 数据的完整性。DaaS 服务提供者不一定可信,至少不可能像企业维护自己的数据库一样可信,因此数据库服务需要保证数据库的内容不会被破坏(数据的完整性)。数据的完整性,由图 1.1 中 DO 处的隐私数据处

理模块(如签名、签名链等)实现,是指 DO 需要提供额外的机制来保证 DSP 对 DR 提交的查询的返回结果是完整的,即返回的查询结果是真实的来自数据拥有者的原始数据,并且没有任何篡改。实际上,DaaS 提供的数据的完整性也存在两层含义:一是保证数据来源的真实性,确实是取自 DO 的数据,这个完整性也称作真实性;二是保护数据不被未授权的人修改,这是通常意义上的完整性。

(3) 数据的完备性。服务提供者不能随意对数据拥有者的数据进行删除、修改或添加自己恶意的数据,因此数据库服务需要保证服务提供者提供的数据是正确的,返回客户的结果是完备的(数据的完备性)。数据的完备性,由图 1.1 中的 DSP 和 DO 共同实现,是指 DO 需要提供额外的机制(如验证结构)来保证 DSP 对 DR 提交的查询结果是完备的,也就是说查询在整个数据库上能够正确执行,并返回所有满足查询条件的元组,DSP 不能恶意地向委托的数据库中添加元组或删除已有元组。保证查询结果的完备性,就是查询结果应该是未经删减过的数据库拥有者实际委托给 DSP 的原始数据(内容和元组个数相同)。

(4) 查询隐私保护。服务提供者不能察觉客户查询相应数据的目的,客户能够从 N 个数据元素中检索第 i 个元素而不被服务提供者发现客户对第 i 个元素感兴趣(查询隐私保护)。查询隐私保护,也称作隐私信息检索,通过图 1.1 中 DR 中的查询转换/结果过滤模块实现,是指 DO 的数据库在委托给 DSP 后,为了保护 DR 的查询意图,DR 需要提供保护请求者隐私的查询,仅通过这个查询,DSP 不能分析请求者的查询目的和操作行为,从而也不能分析 DR 的行为模式。

(5) 访问控制增强。数据库服务在保证上述 4 个方面的安全与隐私的同时,也需要保证数据库的可用性,否则这样的数据库服务没有实际应用价值。所以在数据库服务的可用性方面也提出了挑战:除了通过建立某个属性的索引信息保证数据库的可用性外,还要开发数据库服务中的安全有效的访问控制技术,使得在满足上述安全的情况下实现用户授权访问(访问控制)。为了保护数据隐私和策略隐私,DaaS 中的访问控制策略一般由 DO 定义和维护,这将导致 DO 成为系统中的单点通信瓶颈。为了解决这个问题并有效提高密文数据库的可用性,DO 的访问控制策略需要在服务

提供者端增强(访问控制增强)。1.5节将详细描述数据库服务模式下的访问控制。

1.5 数据库服务模式下的访问控制

根据1.2节描述的数据库服务概念以及1.3节描述的数据库服务系统架构,数据库服务模式下的访问控制可以描述为图1.2。主要包括三类实体:数据拥有者(Data Owners,DO)、数据库服务提供者(Database Service Provider,DSP)和数据消费者(Data Consumers,DCusers),这里DCuser的功能如同图1.1中的DR。由于DSP处于公共云,因此一般假设DSP是不可信或者半可信的(Curious-but-Honest),即他对用户的敏感信息很好奇(Curious),但是会诚实地(Honest)执行必要的功能性操作,如查询优化、授权访问等。为了避免数据隐私泄漏给DSP,大多数方案[1,4~6,8,17~20]采用数据库加密(参见定义2)实现数据隐私保护。

【定义2:数据库加密】 数据库加密就是利用加密技术将明文(Plaintext)数据库转换成(部分)加密数据库(Encrypted DB,也可称作密文数据库(Ciphertext DB)),使得其只能被拥有加密密钥(Encryption Keys)的DCuser访问,而其他任何DCuser都不可访问。

本文中数据库服务模式下的访问控制如图1.2所示(虚线表示相应步骤是可选择的),主要是指基于加密数据库的访问控制增强(参考定义3),根据实际执行访问授权的实体类型,可将其分为如下三类:DO访问控制增强,DSP访问控制增强,DO-DSP访问控制增强。

【定义3:访问控制增强】 根据传统的访问控制策略如DAC、RBAC、MAC等进行第一层访问授权控制可以保证数据的机密性,但是不能保证数据的隐私,因为明文数据对DSP是可见的。在DaaS模式下,为了进一步保证数据隐私(数据对DSP不可见),通过控制密文数据库的加密密钥的分发进行第二层的访问控制授权,即访问控制增强。

(1) DO访问控制增强:DO首先根据自己的访问控制策略加密数据库,然后将加密数据库(Encrypted DB)委托给DSP管理和维护。在此类访问控制模式中,DO进行访问授权的直接控制,如通过图1.2虚线表示的直

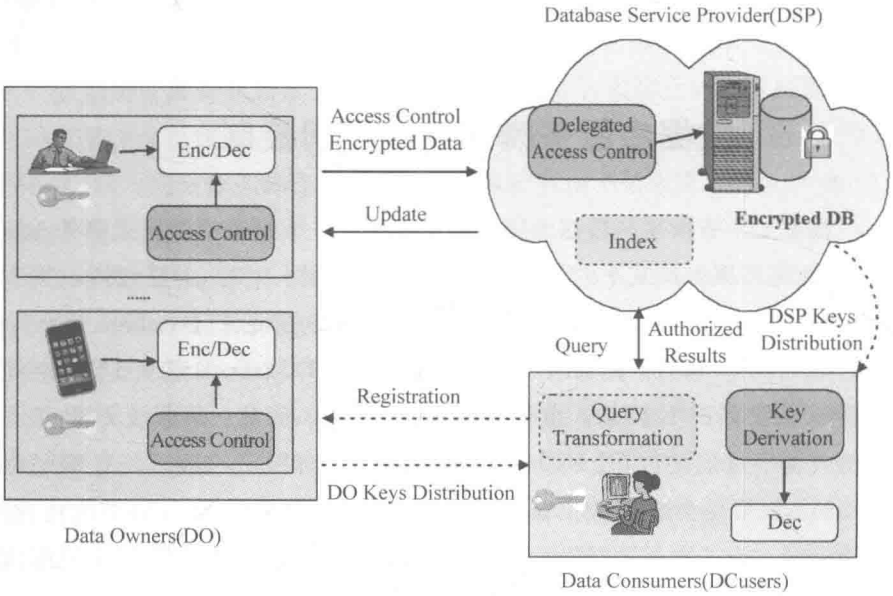

图 1.2 数据库服务模式下的访问控制

接发生在 DO 和授权的 DCusers 之间的密钥分发(DO Keys Distribution)操作实现访问授权控制,然而,DSP 对加密数据库的加密密钥一无所知。

(2) DSP 访问控制增强:DO 不直接进行访问授权的控制,而是将其加密数据库的加密密钥委托给 DSP 管理和分发。在此类访问控制模式中,DSP 可以代表 DO 利益通过图 1.2 虚线表示的直接发生在 DSP 和授权的 DCusers 之间的密钥分发(DSP Keys Distribution)操作实现间接访问授权控制,DSP 知道所有的数据加密密钥。

(3) DO-DSP 访问控制增强:融合了以上两种方案的优点,同时避免了其缺点,如一方面避免了 DO 访问控制增强中 DO 成为通信瓶颈的问题,另一方面避免了 DSP 访问控制增强中 DSP 通过委托的加密密钥窥探 DO 委托的敏感数据的问题。在此类访问控制模式中,DO 需要做两个方面的工作,一是根据自己的访问控制策略选择加密密钥并加密数据库,二是采用一定的访问控制结构(如基于用户属性的访问控制树)保护加密密钥。加密数据库和访问控制结构都委托给 DSP 管理和维护(如图 1.2 中 DO 和 DSP 之间的信息流如 Access Control 和 Encrypted Data 传输)。实际上,

根据对数据消费者 DCusers 隐私保护的程度，这种访问控制增强分为两种：DO-DSP 访问控制增强和保护隐私的 DO-DSP 访问控制增强。

1.6 基于密码学的访问控制增强

根据 1.2 节和 1.5 节的描述知道，要实现 DaaS 模式下基于密码学的访问控制增强，需要考虑以下三个方面：加密方法，访问控制策略以及访问控制策略和加密机制融合。

1.6.1 加密方法

加密方法取决于以下三个方面：加密机制，加密粒度和加密层次。

(1) 加密机制[2,3,20]。加密数据库数据的加密机制如对称加密(Symmetric Encryption，SE)、非对称加密(Asymmetric Encryption，ASE)、代理加密(Proxy Re-encryption，PRE)、基于属性的加密(Attribute-Based Encryption，ABE)、谓词加密(Predicate Encryption，PE)等。

(2) 加密粒度[4,7]。加密数据库数据的加密粒度如文件级(File-grained)、表级(Table-grained)、元组级(Tuple-grained)、属性级(Attribute-grained)。

(3) 加密层次[4](如图 1.3 所示)。加密数据库数据的加密层次如存储层加密(Storage-Level Encryption)、数据库层加密(Database-Level Encryption)、应用层加密(Application-Level Encryption)。

一些著名的安全公司如 RSA[12]、Safenet[13]等提供了数据库加密技术指导，另一方面，一些专业的数据库提供商如 Microsoft SQL Server[9]、Oracle[10,11]、Sybase[14]、IBM DB2[15]等提供了不同粒度的数据库加密机制以及配套的密钥管理机制。加密层次不同，保护的数据安全也有所不同[4]，如图 1.3 所示。存储层加密可以保护由于硬盘的丢失而造成的数据丢失风险，但是存储层加密方法不能实现根据用户权限的选择加密，而且加密粒度是文件。数据库层加密可以在不更改应用的情况下有效支持不同粒度的加密，如表级、元组级和属性级，但是可能会导致 DBMS 性能下降，因为数据库层加密可能导致索引的不可用，除非采用专门的加密算法如保持