基于 ABE 算法的数据库隐私数据加密方法研究

贾 俊[1] 谭振华[1]
JIA Jun TAN Zhenhua

摘 要 针对现有加密方法在对数据库隐私数据加密时，未充分考虑属性基安全性，并且存在关键字、密文索引过大的问题，导致数据库开销较大、安全性不佳，研究了基于 ABE 算法的数据库隐私数据加密方法。为了实现对加密列中数据的灵活查询、降低开销，设计数据库数据外部索引方式，引入 ABE 算法，确定双线性关系，根据列名在十字链表中定位到对应的列，然后依次遍历该列的所有元素，并使用提供的密钥加密处理每个元素，使用相同的密钥解密每个元素，解密后的数据替换原有加密数据，保存在原位置，完成初步隐私数据加密。为进一步增强隐私数据的安全性，通过加密访问控制数据库属性基，优化数据库隐私数据加密方法。通过对比实验证明，新的加密方法应用下，加密后在保证数据安全的前提条件下，有效降低关键词和密文索引数据量，降低数据库计算开销和存储开销，数据量降到了 1.23 Mbit，隐私数据量降到了 18.56 Mbit，促进数据库运行性能提升。

关键词 ABE 算法；隐私；加密；数据；数据库

doi：10.3969/j.issn.1672-9528.2024.10.044

0 引言

数据库作为信息存储和管理的核心载体，其安全性与隐私保护问题日益凸显。数据库隐私数据的加密方法研究，对于保护用户隐私、维护信息安全具有重要意义。当前，数据库加密技术已取得了长足进步，然而现有加密方法仍存在诸多不足。首先，对称加密算法虽然速度快，但密钥管理困难，一旦密钥泄露，整个加密数据库将面临严重威胁。哈希算法虽然可以实现数据的单向加密，但一旦数据被篡改，其完整性将无法得到保障。针对该问题，相关学者进行了研究，如黎明[1]以支持快速查询为出发点，研究了嵌入式数据库加密技术，但是该方法存在开销大的问题。裴江艳[2]以 MD5 数据库加密为核心，研究了信息数据安全存储方法，但是该方

1. 晋中信息学院大数据学院、信息工程学院 山西晋中 030800
[基金项目] 晋中信息学院校级教改项目的课题“基于课程思政视域下网络安全课程教学方法的探索与研究”

Artificial Intelligence Conference,[Volume 1 of 3].Piscataway, N.J.: IEEE, 2019: 1717-1720.
[3]ABDULSADA H F, MOHAMMED S J.Comparison of bit error rate and performance analysis for multicode CDMA techniques in fading and AWGN channels[C]//2022 5th International Conference on Engineering Technology and its Applications. Piscataway: IEEE, 2022:212-217.
[4] 李召飞 . 综合测控终端码分多址干扰分析 [J]. 数字通信世界 ,2024(3):45-47.
[5] 曹建勋 , 王谋业 , 杨胜辉 , 等 . 全双工辅助的无人机上行通信抗窃听物理层安全技术研究 [J]. 现代电子技术 , 2024, 47(9): 21-28.
[6] 雷维嘉 , 毕文佳 , 雷宏江 , 等 . 时间反转 OFDM 系统中增强安全性能的功率分配与人工噪声设计 [J]. 电子学报 , 2024, 52(5): 1570-1581.
[7] 梁丽芳 , 杜小妮 , 李锴彬 , 等 . 基于 Feistel 结构的分组密码算法 Eslice[J]. 山东大学学报 (理学版),2023,58(3):85-92.
[8] 黄琪 , 杨宇晓 , 江陈卓 . 组合跳变随机平移宽间隔混沌跳频序列设计 [J]. 电讯技术 , 2022,62(6):755-761.
[9] 王超 , 温涛 , 段冉阳 .NIST 随机性检测方法研究 [J]. 信息技术与网络安全 , 2018, 37(11):5-8+15.
[10]LEMPEL A, ZIV J.On the complexity of finite sequences[J]. IEEE transactions on information theory, 1976,22(1):75-81.

【作者简介】
曾梦姣（2000—），女，湖北荆门人，硕士研究生，研究方向：电力网络安全。

（收稿日期：2024-07-12）

法随着数据量的增加，其开销也较大。熊爱明等人[3]提出了一种基于混沌映射的数据库信息隐私加密储存算法，该方法通过混沌映射加密了隐私信息，但是其在应用过程中，由于未充分考虑开销问题，导致加密后的数据开销较大。赵中军等人[4]研究了基于AES加密算法的数据库透明加解密系统，其以AES加密算法为核心，设计了一个加密系统，但是系统应用后，其加密的开销较大，导致应用效果不佳。

因此，本文旨在深入研究基于ABE算法的数据库隐私数据加密方法，探索更为高效、安全的加密技术，并且降低开销，以满足不同安全需求的数据库加密需求。

1 数据库数据外部索引方式设计

针对关键字、密文索引过大，对数据库计算开销和存储开销带来负面影响的问题，设计数据库数据外部索引方式。为了实现加密列中数据的灵活查询，数据库专门针对每个经加密的数据列建立对应的索引文档。这些索引文件是由大量的索引记录组成的，每个记录都含有两个重要的信息，即从加密行的原始明文中抽取的检索代码，以及在数据库中与之匹配的索引字段的内容，见图1。

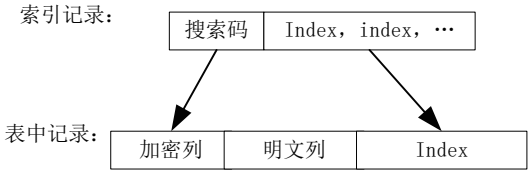

图1 索引记录内容

该方法采用稠密格式，即在已加密的数据列中，每个被加密的数据列都会被指定到与之相对应的索引，这样就保证了查询的精确与高效。在建立密码学的索引数据时[5]，首先从资讯中撷取搜寻码。在该过程中，数据库中的数据存在两种形式：一种是数值型数据，一种是字符型数据。数值型数据因其固有特性，可直接使用其原始数据作为搜索码。然而，对于字符型数据，处理方式则相对复杂。字符型数据涵盖中文、英文以及其他语种。英文数据中的字与字之间有空格隔开，因此可以直接抽取出每一个字作为搜寻代码。而中文语料，为了提高检索效率，需要采用中文自动分割技术，将原始文本分割成若干小块。在充分大的机器字典中，根据某种策略，将要被分析的汉字序列与其对应。对于由 i 个基元（基元可以是字、词、短语等）组成的句子可表示为：$s=w_1,w_2,\cdots,w_i$，则其概率计算公式为：

$$p(s)=p(w_1)p(w_2|w_1)p(w_3|w_1w_2)\cdots p(w_i|w_1w_2\cdots w_{i-1}) \quad (1)$$

式中：w_i 表示构成句子的一个词，每个词 w_i 的出现概率都与其前面的词有关。随着历史长度的增加，不同的历史数目会呈指数增长，这使得直接计算所有可能历史的概率变得不实际。

2 基于ABE算法的隐私数据加密

在数据库数据外部索引方式的基础上，引入ABE算法加密数据库隐私数据，利用用户的属性来决定对数据的访问权限。ABE算法在密文中嵌入政策，在密钥中嵌入属性。在ABE系统中，根据使用者的属性决定对其进行访问。具体来说，ABE将用户的属性和访问策略映射到一个访问结构中，并根据访问结构进行数据的访问控制，在加密的同时，解决关键字、密文索引过大的问题。如果使用者的属性符合存取原则，则使用者可解密资料。ABE的加密流程是：将数据的全部属性映射为一棵存取树；在此基础上，提出了一种新的方法，即加密者依据存取策略，选取一个随机数作为密匙，利用ABE算法加密处理该随机数，并将其与加密后的数据一同发送到用户手中。在解密时，使用者用自己的私有密钥及公开密钥来解密密码。私有密钥的产生依赖于用户的属性、主密钥和公开参数，而公共密钥是公开的。在ABE中完成双线性陪读操作，假设两个阶为 p 的乘法循环群 G 和 GT，在 G 中获取一个生成元 g，用 e 表示 G 和 GT 之间的双线性配对。双线性配对满足以下属性：首先，双线性：对于所有的元素 a 和 b 均属于 Z_p，g 和 h 均属于 G，且存在下述等式关系：

$$e(g^a,h^b)=e(g,h)^{ab} \quad (2)$$

其次，存在非退化性，即 $X\in G$，且 $e(g,g)^1$。加密和解密通过两个接口函数实现，即EncryptDatas()用于数据加密，而DecryptDatas()用于数据解密。这两个函数的设计旨在简化数据库表中特定列数据的加解密过程。加密函数输入参数包括函数接收列名、密钥和指向十字链表首节点的内存指针作为输入。它首先根据列名在十字链表中定位到对应的列，然后依次遍历该列的所有元素，并使用提供的密钥加密处理每个元素。加密后的数据将替换原有数据[6-7]，保存在原位置。加密后的元素可表示为：

$$\text{encryptedData}(i)=E(\text{data}(i),\text{key}) \quad (3)$$

式中：encryptedData(i) 表示加密后的元素；data(i) 表示每个元素；E 表示加密算法；key表示密钥。

解密函数输入参数包括：列名（columnName）、密钥（key）、内存指针（memoryPointer）。与加密函数类似，解密函数也根据列名在十字链表中定位到对应的列。然后，它遍历该列的所有加密元素，并使用相同的密钥解密每个元素。解密后的数据将替换原有加密数据，保存在原位置。解密后的元素可表示为：

$$\text{data}(i)=D(\text{encryptedData}(i),\text{key}) \quad (4)$$

式中：data(i) 表示解密后的元素；D 表示解密算法。在访问数据库表时，将数据列存储在存储器中，并将其作为一个十

字链。这种数据结构允许高效地访问和修改表中的数据。当需要加密或解密特定列时[8-9]，数据加密模块通过提供的内存指针直接定位到十字链表的起始位置，然后从字段名称中发现相应的数据列。密码和解密功能将遍历该数据列中的所有元素，执行相应的加密或解密操作[10]，完成隐私数据初步加密。

考虑到隐私数据本身已经被加密，但如果没有对属性基进行加密，攻击者仍可能通过分析或猜测属性信息来推断出数据的某些特性或模式的问题，在完成隐私数据加密后，提出了一种基于属性基的访问控制方法，进一步提高数据库隐私数据的安全性。数据库将用户的角色划分为两个类型，一种是管理员用户，一种是一般用户。系数据库管理员用户负责管理一般用户账号，分配属性集合，调整用户属性集合，建立权限控制规则等。其不仅负责用户账户的创建与维护，还需确保访问控制策略的有效实施，从而保障数据库的安全性和数据的机密性。

普通用户则持有与自身属性集相对应的密钥，这些密钥是用户访问授权范围内加密数据库信息的必要条件。此外，用户还可以根据需求选择已存在的访问控制规则树来加密信息，以满足特定的安全需求。

在属性基加密方面，利用这种加密方式为普通用户提供精细化的访问权限控制。基于 ABE 算法为密码体制建立了一套授权集，只有在满足条件的情况下，ABE 算法有效解密密码体制。这就意味着加密程序只需要考虑密码的存取规则，而不需要考虑解密密钥的传输方式。

加密中间件扮演着可信第三方的关键角色。当数据库管理员建立新的一般使用者时，会指派一组特殊的属性给使用者，而且对应的私密金钥是基于这个设定而产生的。这些私有密钥和其他有关使用者的数据，都被安全地储存在中间密钥数据库的数据词典内，详细的储存格式见表 1。

表 1 存储格式

序号	用户名	用户私钥
（1）	User1	Privatekey1
（2）	User2	Privatekey2
（3）	……	……
（4）	User*n*	Privatekey*n*

在需要更改一般使用者的数据存取权限时，管理员从数据词典中撷取使用者的数据，即可设定使用者的新属性。该算法根据新的属性设置，自动产生一个新的私有密钥，取代原来的私有密钥。通过该方式可以在保证数据安全、完整的情况下动态调整，保证用户访问权限。

授权管理员使用者设定存取控制规则，既可减轻一般使用者使用密码时的作业负荷，又可以统一管理使用者信息等。对于特定的应用程序，数据库管理员可以设定不同的安全级别并分级。所建立的存取规则会被安全地储存在数据库数据词典中，以供日后的使用与管理。这样的设计使得权限管理更为高效和便捷，同时提升了数据库的安全性和灵活性。

3 对比实验

3.1 实验准备

为进一步验证上述基于 ABE 算法的加密方法在实际应用中的性能，以某数据库为实验研究对象，在数据库中加密保密数据。同时，将该方案与文献 [1] 方法和文献 [2] 方法进行比较，以确保试验结果的可比性。为方便论述，将本文上述提出方法设置为实验组，将文献 [1] 方法和文献 [2] 方法分别设置为对照 A 组和对照 B 组。

实验环境为：开发所用工具为 VC 与 Parser Generator，测试数据库则选用 SQL Server 2005。实现数据库连接时，利用 ODBC 驱动在内网中进行远程连接，确保数据传输的稳定与高效。本地连接速度高达 100 Mbit/s，充分满足对数据交换的需求。运行环境选用配备 Intel Celeron CPU 450 @ 2.20 GHz 处理器、2 GB DDRII 内存、160 GB SATA-I 硬盘以及 Wndows XP SP3 操作系统的主机，以确保流畅运行。数据库运行所需的主机，则配置为 Intel Celeron CPU 430 @ 1.80 GHz、同样拥有 2 GB DDRII 内存、160 GB SATA-I 硬盘以及 Windows XP SP3 操作系统的硬件配置，保证数据库的稳定运行与高效处理。量化三种加密方法的应用性能，通过对比加密后关键字大小和密文索引大小，实现对三种方法计算开销和存储开销的对比。关键字大小和密文索引越大，计算开销和存储开销越大，则说明加密方法性能越差；反之，关键字大小和密文索引越小，计算开销和存储开销越小，则说明加密方法性能越强。将上述 SQL Server 2005 数据库中的数据随机分为 5 组，每组数据量和隐私数据量统计如表 2 所示。

表 2 数据库数据分组情况记录表

序号	分组	数据量 /Mbit	隐私数据量 /Mbit
（1）	第一组	1000	100
（2）	第二组	2000	200
（3）	第三组	3000	300
（4）	第四组	4000	400
（5）	第五组	5000	500

3.2 结果分析

针对上述五组数据，根据上述论述，将得到的结果记录如表 3 所示。

表 3 三种加密方法加密效果对比表

单位：Mbit

组别	实验组		对照 A 组		对照 B 组	
	关键字大小	密文索引大小	关键字大小	密文索引大小	关键字大小	密文索引大小
第一组	0.45	10.25	1.53	20.36	5.36	25.26
第二组	0.46	12.26	2.03	24.56	8.36	29.56
第三组	0.81	14.25	2.65	28.36	11.26	33.56
第四组	1.03	16.52	3.15	32.46	14.56	37.65
第五组	1.23	18.56	3.85	40.25	18.65	45.65

结合表 2 和表 3 中详细罗列的数据，可以清晰地观察到一种趋势：随着数据组中数据量和隐私数据量的不断增加，加密处理数据时，无论是关键字的大小，还是密文索引的大小，均在一定程度上呈现出增加的态势。这种增加并不是无序的，而是遵循着一定的规律，显示出加密过程中数据处理的复杂性和精细性。

进一步通过横向对比不同加密方法的效果，可以发现，实验组的加密方法在各方面都表现得尤为出色。具体来说，在使用实验组的加密方法加密每组数据后，其关键词的大小和密文索引的大小均小于三种方法，在数据量和隐私数据量最高的第五组实验时，其数据量降到了 1.23 Mbit，隐私数据量降到了 18.56 Mbit，数据量比对比方法降低了 2 Mbit 以上，隐私数据量比对比方法降低了 20 Mbit 以上。这不仅意味着在加密过程中，实验组方法能够更加高效地处理数据，减少不必要的数据膨胀，还意味着加密后的数据占用更少的存储空间，更加节省资源。

更为关键的是，从增长趋势来看，实验组加密方法的关键字大小和密文索引大小的增长趋势也最为平缓。这意味着即使数据量进一步增加，实验组加密方法依然能够保持相对稳定和高效的性能，而不会像其他方法那样出现急剧增长。这种稳定性在实际应用中具有非常重要的意义，尤其是在处理大规模数据时，能够显著减少数据库资源的消耗和成本的投入。

4 结语

数据库隐私数据加密方法的研究，是信息安全领域的重要课题。本文通过引入 ABE 算法，提出了一种全新的加密方法。然而，数据库加密技术的研究并非一蹴而就，仍需不断探索和创新。未来，在这个技术飞速发展的时代，云计算将进一步普及和深化，大数据处理能力也将显著提升，人工智能的应用范围将更加广泛，数据库加密将面临更为复杂和严峻的挑战。因此，需要持续关注数据库加密技术的发展动态，加强与国际先进技术的交流与合作，共同推动数据库加密技术的进步，为信息安全事业贡献力量。数据库加密只是信息安全防护的一部分，还需要与其他安全措施相结合，形成多层次、全方位的安全防护体系。例如，加强数据库的安全审计、访问控制、漏洞管理等，都是提升数据库安全性的重要手段。只有综合运用各种安全技术和管理措施，才能有效保护数据库的隐私数据，确保信息的安全与完整。

参考文献：

[1] 黎明 . 基于支持快速查询的嵌入式数据库加密技术研究 [J]. 浙江水利水电学院学报 , 2023,35(6):86-90.

[2] 裴江艳 . 基于 MD5 数据库加密的企业档案信息数据安全存储方法 [J]. 信息与电脑 (理论版),2023,35 (11): 242-244.

[3] 熊爱明 , 李明倩 , 刘芳 . 基于混沌映射的数据库信息隐私加密储存算法 [J]. 吉林大学学报 (信息科学版), 2023, 41(3): 459-464.

[4] 赵中军 , 杨阳 , 杨兴 , 等 . 基于 AES 加密算法的数据库透明加解密系统的设计与实现 [J]. 通信技术 , 2023,56(3):377-382.

[5] 高永相 , 崔长杰 , 李宇航 . 嵌入式 Linux 系统 SQLite 数据库加密功能的移植和应用开发 [J]. 工业控制计算机 , 2023, 36(3): 25-26+29.

[6] 杜璞 , 张小艳 . 基于面向对象技术的数据库外层敏感信息加密系统设计 [J]. 自动化与仪器仪表 ,2023(3):245-248+259.

[7] 段晓聪 . 基于无线传感网络分簇策略的分布式数据库加密存储研究 [J]. 传感技术学报 ,2022,35 (12):1728-1732.

[8] 林加华 , 李志虹 , 姜华 . 一种基于安全散列加密算法的数据库操作痕迹可证明追踪算法 [J]. 现代计算机 , 2022, 28(22): 42-46.

[9] 张黎 , 骆春山 , 谢委员 , 等 . 基于分支混淆算法的隐私数据库自适应加密方法 [J]. 计算机与现代化 ,2022(3):43-47.

[10] 李骥才 , 刘丹 . 基于位置匿名的数据库隐私可搜索加密算法 [J]. 计算机仿真 ,2022,39(1):423-426+431.

【作者简介】

贾俊（1981—），男，山西晋中人，硕士，讲师，研究方向：计算机网络、网络安全。

谭振华（1998—），男，山西朔州人，硕士，助教，研究方向：农业物联网。

（收稿日期：2024-06-12）