分区加密与结构化加密相结合实现 OLAP 的数据安全性

武 彤,漆 媛
(贵州大学 计算机科学与技术学院,贵州 贵阳 550025)

摘 要:随着数据仓库、OLAP、数据挖掘技术的普及,基于数据仓库的决策支持系统在企业得到了广泛应用。而企业决策支持系统所涉及的数据是企业决策所需的重要数据,这些数据的丢失或泄露将直接影响企业的生存与发展,确保企业决策系统 OLAP 的安全性从而保护数据安全有其重要的意义。因此 OLAP 系统的数据安全性已成为数据安全领域的研究热点。以某"生产线质量控制决策分析系统"为研究平台,分析了系统的安全需求,针对平台上 OLAP 系统的多维数据,结合分区加密及结构化加密的特点,使用分区加密与结构化加密相结合,实现了 OLAP 系统的数据加密。通过实践证明,这种混合加密策略既能保证数据的安全性,又不影响 OLAP 系统的性能。该研究方案在同类系统中具有一定的推广应用价值。

关键词:数据仓库;OLAP;数据加密;分区加密;结构化加密
中图分类号:TP309.2 **文献标识码**:A **文章编号**:1673-629X(2017)01-0089-04
doi:10.3969/j.issn.1673-629X.2017.01.020

Implementation of Data Security of OLAP System Combined Partition Encryption with Structured Encryption

WU Tong, QI Yuan
(College of Computer Science & Technology, Guizhou University, Guiyang 550025, China)

Abstract: With the popularization of data warehouse, OLAP and data mining technology, decision support system based on data warehouse has been widely used in enterprise, and the data involved in enterprise decision support system is important for enterprise decision-making. The loss or disclosure of these data will directly affect the survival and development of enterprises. To ensure the security of the enterprise decision-making system OLAP in order to protect data security has its important significance. Therefore, the data security of OLAP system has become a research hotspot in data security field. The security requirements of the system is analyzed based on a "quality control decision analysis system of production line". According to the multi-dimensional data of the OLAP system on the platform, combined with the characteristics of partition encryption and structured encryption, the combination of them is used to achieve the OLAP system data encryption. The practice shows that this hybrid encryption strategy can not only guarantee the security of data, but also does not affect the performance of OLAP system. The research program has a certain value in the same kind of system.

Key words: data warehouse; OLAP; data encryption; partition encryption; structured encryption

0 引 言

随着数据库技术的深入发展及普及应用,企业已经在不同的时期针对具体应用建立了大量的数据库应用系统,这些数据库应用系统解决了企业在生产、管理中的很多问题,同时推动了数据库技术的发展。但是面对不断增加的数据,企业用户不再满足于数据库应用系统提供的功能,而且提出了深层次的应用问题:能不能从数据中提取信息或者知识为决策服务,就数据库技术而言是无能为力的,由此产生了数据仓库、OLAP(联机分析处理)以及数据挖掘技术。随着数据仓库、OLAP 以及数据挖掘技术的发展,基于数据仓库的决策支持系统在企业得到了广泛应用[1]。

数据仓库是一个综合的解决方案,主要帮助企业有关主管部门和业务人员做出更符合企业发展规律的

收稿日期:2015-11-17 **修回日期**:2016-02-25 **网络出版时间**:2017-01-04
基金项目:贵州省科技计划工业攻关项目(黔科合 GY 字[2010]3061)
作者简介:武 彤(1964-),女,硕士,教授,CCF 会员,研究方向为数据仓库;漆 媛(1989-),女,硕士研究生,研究方向为数据库技术。

决策。而 OLAP、数据挖掘与数据仓库的协同工作,一方面可以迎合和简化决策分析过程中的重要步骤,提高决策分析的效率和能力,确保决策分析中数据来源的广泛性和完整性。另一方面,OLAP、数据挖掘技术已经成为数据仓库应用中极为重要和相对独立的方面[2]。OLAP 是使分析人员、管理人员或执行人员能够从多角度对信息进行快速、一致、交互地存取,从而获得对数据的更深入了解的一类软件技术。数据仓库中存储的巨大数据量能够科学地反映历史信息,通过 OLAP 技术对数据进行处理与多维分析,可以帮助企业管理者从数据中获得支持决策的信息,从而做出正确的决策。这对帮助企业提高市场竞争力,完善企业管理,针对性地制定出相关政策意义重大。由此可以看出,OLAP 分析所涉及的数据是决策所需的重要数据,这些数据的丢失或泄露将直接影响企业的生存与发展,因此确保企业决策系统 OLAP 的安全性从而保护数据安全有其重要的意义。文中以某企业生产线质量控制决策分析系统为研究平台,研究以分区加密和结构化加密相结合实现 OLAP 的安全性。

1 分区加密与结构化加密

OLAP 系统的数据来源于数据仓库,确保数据仓库中数据安全的最有效方法就是数据加密。对数据仓库内部的数据进行加密时,常用的方法有分区加密和结构化加密。

(1)分区加密。

分区加密就是对数据仓库的不同部分应用不同的加密模块。如针对一个数据仓库,可以根据数据重要程度的不同,将数据划分成 A,B,C,…不同的分区,对不同的分区数据采用不同的加密算法进行加密,当用某一分区的编码器对另一分区的数据进行解密时,将得到一堆垃圾。即不同的加密/解密过程可以有效地分到不同的分区中,以保证数据的安全性。

(2)结构化加密。

在数据仓库中,数据的内部结构和数据的类型在加密/解密过程中必须被保护。例如在一个非加密状态,如果域 A 在域 B 之前,域 B 又在域 C 之前,那么在加密状态下,也应该维护这样的结构。如果一个 ASCII 域被加密,那么加密值也必须是 ASCII。在加密过程中维护域是一个非常复杂的问题,并且加密数据使得像汇总之类的处理很难进行,因此需要选择性地对数据仓库中的数据进行加密[3]。通常只有非键值,非索引的数据被加密。加密键值和索引使得本来复杂的过程更加复杂,还有一些其他的问题要考虑。在决定哪些数据加密时,最简单的方法就是把一个表中的所有非键值/非索引值加密。一种情况是只加密一列数据,所有其他的数据都以原始的状态存储。另外一种情况是加密数据的特定行。还有一种情况,加密算法是周期性变化的。因此请求者不仅要使解密算法和加密数据相匹配,而且这个解密算法在这个时刻是正确的[4]。

2 OLAP 系统的多维分析及安全需求

文中研究内容基于的平台是某企业“生产线质量控制决策分析系统”,该系统将对生产线上的电视机产品质量围绕“基板品质”及“电视机质量”两个主题进行 OLAP 分析。

2.1 基板品质分析

“基板品质”分析是通过电视机主板生产环节的各个质量信息采集点的数据处理汇总及分析,可以为决策者们提供事实数据的有力支持,因此,决策分析系统将基板品质分析作为主要分析主题是非常必要的。基板分析主题主要以基板的合格数量、出错数量来计算基板品质的合格率等,并将其作为基板品质的度量信息。

有如下分析需求:

(1)某段时间内,某种基板类型的合格率。

欲得到查询结果,需要连接版型维度表、时间维度表以及版型分析事实表三张表。

(2)某段时间内,某个生产线上工作生产的基板合格率。

欲得到查询结果,需要连接生产线维表、时间维度表以及生产线分析事实表三张表。

(3)某段时间内,某个采集点上的基板合格率。

欲得到查询结果,需要连接采集点维度表、时间维度表以及基板类型分析事实表三张表。

(4)某段时间内,生产线上工作生产的某种版型的基板合格率。

欲得到查询结果,需要连接版型维度表、时间维度表、生产线维表、生产线分析事实表四张表。

(5)某段时间内,生产线上工作生产的某种版型的基板合格率。

欲得到查询结果,需要连接版型维度表、时间维度表、采集点维度表、版型分析事实表四张表。

(6)某段时间内,产生某种故障现象的基板数量。

欲得到查询结果,需要连接故障现象维表、时间维度表、版型故障现象分析事实表三张表。

(7)某段时间内,产生某种故障类型的基板数量。

欲得到查询结果,需要连接故障现象维表、时间维度表、版型故障现象分析事实表三张表。

(8)某段时间内,某采集点上,因何种故障原因所致的故障数量。

欲得到查询结果,需要连接采集点维度表、故障现象维表、时间维度表、版型故障现象分析事实表四张表[5]。

从以上查询内容可见,企业用户所关注的查询大多需要连接 3 张或以上的维表与事实表。

2.2　电视机质量分析

电视机质量分析主题中,是对主板生产之后一直到电视机成品下线整个生产环节中产品的质量控制数据进行多维分析。包括故障数量、产品直通率等,最终以分析数据说明质量控制情况[6]。该主题分析与基板品质主题分析一样,按时间粒度(日,月,季,年)进行多角度的切片、切块、钻取分析。具体分析需求不再赘述。

2.3　OLAP 系统的安全需求

该 OLAP 系统中的数据是从数据仓库的海量历史数据中提取出的具有针对性的决策型数据,以构建 OLAP 多维分析模型。因此,数据加密是为了防止非法用户进入 OLAP 系统后,盗取重要的企业决策数据。由于数据在数据仓库中的结构、数据类型等在数据进行加/解密过程中都不能有所改变,一旦进行加/解密过程,都将占用一定的机器资源,从而对 OLAP 系统分析效率造成一定的影响。因此,需要选择既能保证数据安全性,又对 OLAP 系统性能影响最小的数据加密安全方案[7]。选择的加解密方法必须满足如下要求:

(1)安全性:安全机制须能防止非法访问数据操作,预防恶意推理。

(2)适用性:安全机制应适用于本生产线质量控制决策分析系统,并能在不需要重大修改的情况下应用在多种环境中。

(3)有效性:安全机制须满足 OLAP 系统的在线处理和实时响应的性能要求。该系统中,要求加入安全控制后,系统响应时间不能超过加入安全控制前原响应时间的一半。

(4)可用性:安全机制对于合法用户来说,数据必须是可访问的,只在某些情况下才设置合理的访问约束。

(5)实用性:安全机制不应要求对现存的 OLAP 系统结构做重大修改,而应能够利用现存的查询处理机制和安全机制[8]。

3　OLAP 系统的分区加密

由于“生产线质量控制决策分析系统”是一个实际运用的 OLAP 系统,实施的安全加密方案需满足 OLAP 的可用性与快速决策性等需求,这就决定了所采用的加密算法必须具有安全、加密/解密速度快等特点,并且要在数据存储、实现平台方面具有较强的可塑性。通过实验对比分析后,采用 SHA-1 和 AES 算法进行分区加密。

3.1　分区加密的设计

在该 OLAP 系统中,分为系统用户管理与 OLAP 查询分析两大模块。用户管理模块主要是对 USER 表中数据进行管理,该模块用于管理用户名和密码、角色的分配、权限的控制等。OLAP 查询分析模块是对基板品质、电视机质量分析主题进行数据分析与处理。采用两种不同的加密算法对用户管理模块与 OLAP 查询分析模块所涉及数据进行分区加密。

系统用户管理模块中,采用 SHA-1 加密算法对用户密码进行加密。由于登录密码需要在增加用户时便存入加密后的密文密码,而且在系统管理级别的超级管理员用户对 USER 维表中 Password 审查过程中,SHA-1 所具有的对同一数据加密后密文相同的特点,当出现两者以上用户密文相同的情况时,系统管理级别用户便可提醒密码设置重复的用户更换更为复杂的登录密码以保证密码安全。

OLAP 查询分析模块对加密算法的加解密速率有较高要求。AES 算法由于其加密实现快、加解密效率高的特点[9],用来加密决策分析查询的各维表与事实表中重要数据。首先对参与决策查询分析的各维表与事实表中数据用 AES 加密后重新导入维表/事实表中,当提交查询申请后,系统程序调用 AES 解密算法,对所有表内数据进行解密,然后恢复维表与事实表间连接,从而进行查询分析,并将结果以直观的方式显示。

3.2　分区加密的实现

(1)SHA-1 算法加密用户登录密码。

采用 SHA-1 算法对该 OLAP 系统中用户表(USER)的用户密码进行加密。超级管理员在添加用户时,可为该用户设置原始密码,用户可自行进行密码修改。加密系统对用户修改后的密码实施加密,以密文形式直接存入用户表 Password 列中,系统管理员级别用户(即超级管理员)也无法破解用户的密文密码。由于 SHA-1 算法是单一加密,即相同的加密内容能得到相同的密文字符串。当用户设置密码后,提交进行存储时,对显示相同密文字符串的用户,可提醒更改为更复杂的密码[10]。

(2)AES 算法加密维表与事实表数据。

该 OLAP 系统有两个分析主题,分别是“基板品质”和“电视机质量”,两个分析主题都采用星座模型进行数据的组织与存储。“基板品质”主题的 OLAP 模型如图 1 所示。

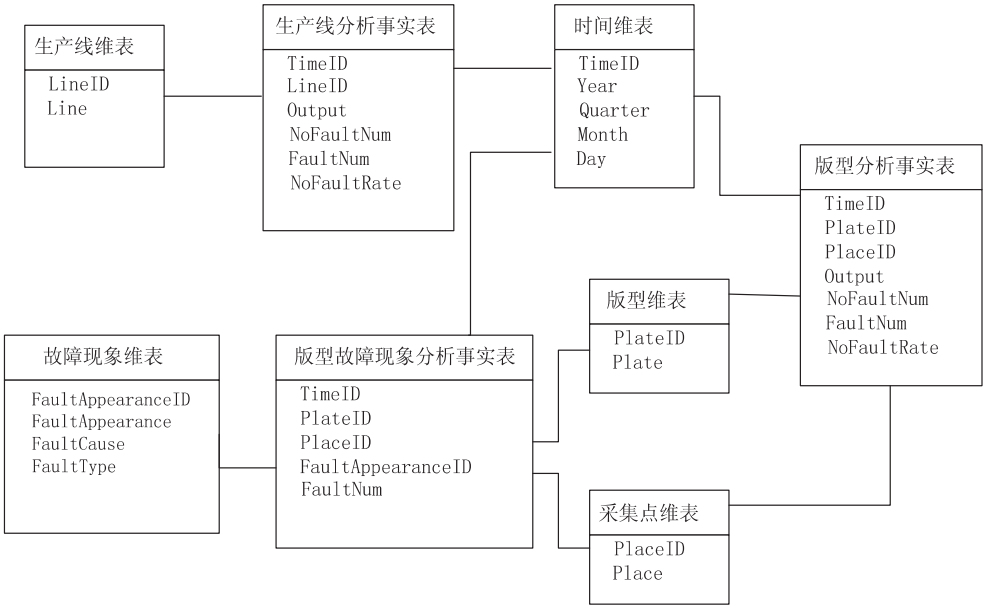

图 1 基板品质主题分析模型

如图所示,基板品质分析模型中有 5 个维表与 3 个事实表,其中事实表之间共享维表,组成了星座模型[11]。电视机质量分析模型与此相同。对于数据仓库中存储的这 8 张表的数据,采用 AES 加密算法进行加密存储。当要进行 OLAP 查询分析时,解密系统对多维表数据全部解密后再进行连接查询。可见查询数据时需要对全部多维数据进行解密后,再由表间主键连接多维表进行查询操作。系统查询响应时间花销主要由查询请求提交时间、数据解密时间、数据连接查询时间、数据反馈时间构成,由于请求提交时间与数据反馈时间是一定的,时间花销的差距可以视为数据解密时间与查询时间的总和差。由于查询需求并非对每个维表与事实表数据都要求参与分析运算,对全部数据整体进行解密的环节可能造成时间开销的浪费,不满足查询条件的数据所消耗的解密时间应该是可以被节约的。由此提出如下的结构化加密方法进行改进。

4 OLAP 系统的结构化加密

结构化加密方法仅对 OLAP 系统中部分维表数据以及事实表中重要的敏感数据进行 AES 加密,而用于连接维表与事实表间的主键不予加密。因此,即使在多维表中数据以密文显示的情况下,也能实现表间的连接。采用结构化加密后,可以通过查询得到视图,对视图内涉及的数据进行解密,解密数据范围比分区加密方法大大减少,进而加快了查询响应时间。当用户提交查询请求时,系统只需要调用参与查询的多维表,通过表间联系,将参与查询的数据进行解密,最后将查询结果反馈用户。这样可以节约系统对所有维表与事实表数据解密后再进行连接查询而产生不必要的时间开销。

当用户提交查询分析请求时,系统通过表间主键连接,将符合查询请求的数据查找出后形成物化视图,解密程序调用 AES 解密算法只对视图涉及部分的数据进行解密,并将分析结果反馈用户。

图 2 是没有进行数据加密和进行结构化加密以及分析加密后 OLAP 系统查询分析响应时间的比较。

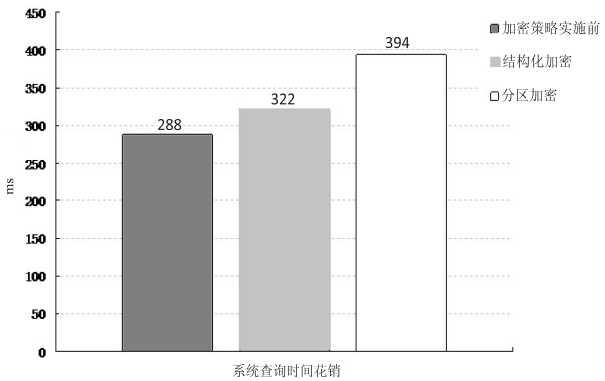

图 2 加密前系统查询与采用加密方案后查询时间花销对比

从图中可以看出,结构化加密作为改进加密方案,节省了系统用于解密后连接数据再进行查询分析的时间花销,采用结构化加密在实现了安全性的同时,降低了对系统查询性能的影响[7]。

5 结束语

文中研究的“生产线质量控制决策分析系统”是一个 OLAP 系统,对该 OLAP 系统中存储的数据,采用不同的加/解密算法对不同的数据进行分区加/解密,如对用户信息采用 SHA-1 算法进行加/解密,对多维数据采用 AES 算法进行加/解密。而对多维数据进行

(下转第 96 页)

DTW 与 VQ 相结合的模型在电子伪装语音存在的情况下,识别性能有很大提升,识别效果明显改善。但在说话人识别领域,该模型的识别效果并不理想,后续的研究可以通过使用改进后的伪装鉴定模型或者选取更为有效的特征参数等方法来进一步提高系统的性能。

3 结束语

电子伪装语音的存在,使得基于 VQ 模型的说话人识别性能降低,识别效果变得不理想。文中利用 DTW 模型匹配出测试语音的伪装程度,再将 VQ 模型训练语音的伪装程度调整至与测试语音同一伪装程度层面,实现对该模型的补偿,使其性能得到明显改善。实验结果表明:经过补偿之后的 VQ 模型对电子伪装语音的识别性能显著提升,识别效果良好。

参考文献:

[1] Neustein A, Patil H A. Forensic speaker recognition: law enforcement and counter - terrorism [M]. [s. l.]: Springer, 2011.

[2] 张翠玲,谭铁军,刘 昇. 伪装语音的自动话者识别研究[J]. 刑事技术,2007(2):18-21.

[3] 张翠玲. 伪装语音的声学研究[D]. 天津:南开大学,2005.

[4] 张桂清,金怡珠,刘红伟,等. 电子伪装语音的变声规律研究[J]. 证据科学,2010,18(4):503-509.

[5] 余建潮,张瑞林. 基于 MFCC 和 LPCC 的说话人识别[J]. 计算机工程与设计,2009,30(5):1189-1191.

[6] Tan T J. The effect of voice disguise on automatic speaker recognition[C]//Proceedings of 3rd international congress on image and signal processing. Yantai: IEEE, 2010: 3538-3541.

[7] Zhang C, Tan T. Voice disguise and automatic speaker recognition[J]. Forensic Sci. Int., 2008, 175(2-3): 118-122.

[8] Wu H J, Wang Y, Huang J W. Blind detection of electronic disguised voice[C]//Proceedings of IEEE international conference on acoustics, speech and signal processing. Vancouver, BC: IEEE, 2013: 3013-3017.

[9] Wu H J, Wang Y, Huang J W. Identification of electronic disguised voices[J]. IEEE Transactions on Information Forensics And Security, 2014, 9(3): 489-500.

[10] 文 翰,黄国顺. 语音识别中 DTW 算法改进研究[J]. 微计算机信息,2010,26(7-1):195-197.

[11] 刘长明,任一峰. 语音识别中 DTW 特征匹配的改进算法研究[J]. 中北大学学报:自然科学版,2006,27(1):37-40.

[12] 丁艳伟,戴玉刚. 基于 VQ 的说话人识别系统[J]. 电脑知识与技术,2008,4(5):1181-1183.

[13] 赵 力. 语音信号处理[M]. 北京:机械工业出版社,2003.

[14] 孔勇平. 矢量量化 LBG 算法的研究[J]. 硅谷,2008(6):39-40.

[15] 王 伟,邓辉文. 基于 MFCC 参数和 VQ 的说话人识别系统[J]. 仪器仪表学报,2006,27:2253-2255.

(上接第 92 页)

加密时,采用了结构化加密的思想,只对 OLAP 系统中部分维表数据以及事实表中重要的敏感数据进行加密,用于连接维表与事实表间的主键不予加密,从而保证了进行 OLAP 分析时,数据加/解密时间开销不会影响 OLAP 系统的查询分析性能。通过实践证明,这种混合的加/解密方案,既保证了 OLAP 系统中数据的安全性,又不影响 OLAP 系统的查询分析性能。该研究方案在同类系统中有一定的推广应用价值。

参考文献:

[1] 陈京民. 数据仓库与数据挖掘技术[M]. 北京:电子工业出版社,2003:12-14.

[2] 李雄飞,杜钦生. 数据仓库与数据挖掘[M]. 北京:机械工业出版社,2013:18-19.

[3] 王丽珍,周丽华. 数据仓库与数据挖掘原理及应用[M]. 北京:科学出版社,2005.

[4] 于醒兵. 数据库结构加密理论与技术[D]. 秦皇岛:燕山大学,2006.

[5] 李 华. OLAP 技术在生产线质量控制决策系统中的应用[D]. 贵州:贵州大学,2012.

[6] 武 彤,程 辉. 基于决策树算法的电视机故障维修模型设计[J]. 计算机技术与发展,2014,24(5):150-152.

[7] 漆 媛,武 彤. 基于生产线质量控制系统的 OLAP 安全性研究[J]. 计算机技术与发展,2014,24(9):179-182.

[8] 汤姆森. OLAP 解决方案-创建多维信息系统[M]. 朱建秋,译. 第 2 版. 北京:电子工业出版社,2004:8-10.

[9] NIST. Adwanced Encryption Standard (AES)[M]. [s. l.]: Federal Information Processing Standards Publication, 2001.

[10] JavaScript 实现 SHA-1 加密算法的方法[EB/OL]. 2014. http://www.jb51.net/article/62035.htm

[11] 武 彤. 电视机生产线质量控制决策系统 OLAP 模型设计[J]. 微计算机信息,2012(11):283-284.