多级安全数据库系统的推理分析*)

杨 武[1] 李立新[2] 文俊浩[3]

(重庆工学院 计算机系 重庆 400050)[1] (解放军信息工程大学电子技术学院 郑州 450004)[2]

(重庆大学软件学院 重庆 400044)[3]

摘 要 介绍了多级安全数据库系统的推理问题,及用于推理的信息来源和方法策略;研究并分析了目前在多级安全数据库系统中推理问题的成果:推理的形式化,数据库设计中的推理控制技术,数据库系统运行时的推理控制方法,以及数据级的推理控制方法,并指出了各个方法的局限性。

关键词 数据库系统,安全,推理

Inference Analysis on Multi-Level Security Database System

YANG Wu[1] LI Li-Xin[2] WEN Jun-Hao[3]

(Department of Computer Science, Chongqing Institute of Technology, Chongqing 400050)[1]

(Electronic Technology College, PLA's Information Enging neering University, Zhengzhou 450004)[2]

(Software School, Chongqing University, Chongqing 400044)[3]

Abstract The inference for multi-level security data system and its source and scheme are introduced. And, it's latest achievements are analyzed and studied: the formalization of inference, the control technologies of inference in database design, the control methods of inference during database system running, and the control means of inference for data level. Moreover, the drawbacks of schemes are presented.

Keywords Database system, Security, Inference

1 引言

多级安全数据库管理系统利用 MAC(强制访问控制)阻止高安全级的数据,未经授权泄露给低安全级的用户。推理,即从已知的信息推出新的信息。安全数据库系统中的推理问题是指,如何阻止一个用户推理出未被授权访问的信息。尽管在可信任计算机系统评价准则和可信数据库解释中对推理控制没有要求,但推理问题是数据库安全的一个重要威胁,用户利用已知信息并结合推理,可获得未授权访问的敏感信息和数据。

推理不是数据库管理系统 DBMS 设计和实现的问题,而是由一组特定的数据值和一组特定的逻辑蕴涵规则引起的。另外,支持推理的数据不是由 DBMS 的可信计算基中被恶意植入的代码引起的,而是由合法用户的非恶意操作产生的。

在多级安全数据库中,可用于推理的信息主要包括[1]:源于数据库的统计数据;存储在关系中的数据;包括关系名和属性名在内的元数据;数据的存在和缺失;未保存在数据库中的特定领域知识;经过一段时间之后数据的变化和消失;数据的语义在特定应用域中是已知的,但没有保存在数据库中;其它元数据如约束等。

在多级安全数据库系统中可被用来进行推理的主要策略包括:

(1)根据演绎推理进行的推理,通过古典的和非古典的演绎逻辑推理出新的信息。其中古典的演绎推理基于基本的逻辑规则,非古典的演绎推理包括概率推理、动态逻辑等;

(2)通过归纳推理进行的推理,根据观察到的例子,利用一些特定的规则推理出一些假设;

(3)根据类比进行的推理,可以根据 X 类似 Y,在已知 Y 的特性的情况下可以推理 X 的特性;

(4)存在推理,由某些信息判断某些数据项的存在。例如根据"John lives in Boston",可以推理出"存在一个数据项 John";

(5)统计推理,一个数据项的信息可以通过比较含该数据项的不同集合的统计数据得到。

与数据库系统推理问题相关的一个研究领域是人工智能 AI,AI 的研究目标是从已知的知识推理出新的知识。随着计算机技术的发展,特别是随着数据仓库和数据挖掘技术的发展,由于可用来推理的信息和技术非常丰富,在数据库系统中阻止未经授权的推理是一个非常困难的工作,强大的数据挖掘工具使用推理技术从数据中发现隐含的信息。推理对数据库的安全构成了更加严重的威胁。

多级安全数据库系统的推理涉及到推理的形式化、数据库设计中的推理控制技术、数据库运行时的推理控制方法和数据级的推理控制方法等多方面内容。

2 推理的形式化

推理的形式化主要研究如何定义推理问题,目前主要从经典信息论、数据划分、集合理论和函数依赖等多方面进行研究[2]。

2.1 经典信息论

*)项目资助:重庆市自然科学基金支持项目(7969)、教育部科技重点项目,编号:03115。杨 武 硕士,副教授,主要研究方向:分布式处理,信息安全。

利用经典信息论来描述推理问题，其形式化定义如下：设有两个数据条目 x 和 y，$H(y)$ 表示 y 的不确定性，$Hx(y)$ 表示在特定 x 的条件下 y 的不确定性。则在确定 x 的条件下 y 的不确定性缩小为：

$$INFER(x\rightarrow y)=\frac{H(y)-Hx(y)}{H(y)} \quad (1)$$

其中 $INFER(x\rightarrow y)\geqslant 0$ 且 $INFER(x\rightarrow y)\leqslant 1$。如果 *INFER* 值为 0，则从 x 无法推出 y 的信息；如果 *INFER* 值在 0 和 1 之间，那么给定 x，y 可能存在；如果 *INFER* 值为 1，则给定 x 能够推出 y。

利用经典信息论描述推理问题存在两个缺点，确定 $Hx(y)$ 比较困难；没有考虑式(1)的计算复杂性。

2.2 数据划分

研究[3]认为：对每个用户，数据库中的数据可以划分为两个集合：一个可见的集合与一个不可见的集合，用户只允许访问可见集合的数据，不允许访问不可见集合的数据。另外，用户通过查询可得到数据集合 known。如果不可见集合和 known 集合的交集为空，则不存在推理问题；反之，则存在推理问题。

2.3 集合理论

用集合理论来定义推理问题[4]，假设一个数据库中的每一个数据项均拥有一个安全级别〈密级，范畴〉，安全级别是偏序的。“→”关系定义如下：

令 x 和 y 为两个数据项，如果有可能从 x 推理出 y，则 $x\rightarrow y$ 成立；显然“→”关系是自反的、传递的。如果 $\forall x\in S$，且 $x\rightarrow y$，都有 $y\in S$ 成立，则集合 S 称为推理封闭的。

2.4 函数依赖

函数依赖定义如下：设 R 是一个关系模式，$*$ 表示其属性集合，$X,Y\subseteq *$，当其中任意两个元组 u,v 对应于 X 的那些属性分量的值均相等时，则有 u,v 中对应于 Y 的那些属性分量的值也相等，称 X 函数决定 Y，或 Y 函数依赖于 X。

如果一个函数依赖对于低安全级别的用户是已经知道的，就会产生推理问题[5,6]。表 1 是一个例子，显然，(*a*14，S)对于一个 U 级(公开级别)的用户是不可见的，但如果函数依赖(X1，X2)→(Y1，Y2)成立且为 U 级用户所知，则(*a*14，S)，(*a*13，S)就可以被 U 级用户推断出来。

表 1 函数依赖导致的推理

X1	X2	Y1	Y2	W
a11，U	a12，U	a13，U	a14，U	w1
a21，U	a22，U	a23，U	A24，S	W2
a11，U	a12，U	a13，S	A14，S	W3
…	…	…	…	…

3 数据库设计中的推理控制技术

推理控制的数据库设计技术[7]是为了解决如何确定数据库客体(包括数据、元数据以及约束等)安全级别。对数据库设计者和用户来讲，应该清楚输入到数据库中的数据的安全级别。在理想情况下，为了阻止所有未经授权的信息泄露，应该遵守多级安全数据库基本规则：一个数据项的安全级别应该支配所有影响它的数据的安全级别。

现有的技术主要是在多安全级数据库设计期间，使用安全约束为数据库模式指定适当的安全级别，并将安全约束以语义数据模型进行表示。该方法的缺陷是在实际应用中，提高数据项的安全级别会受到限制。

4 数据库运行时的推理控制方法

运行时的推理控制机制是对数据库设计中推理控制技术的一个补充，运行时的推理控制机制主要包括对查询响应进行修改、多实例、审计三方面[6]。

4.1 对查询响应进行修改

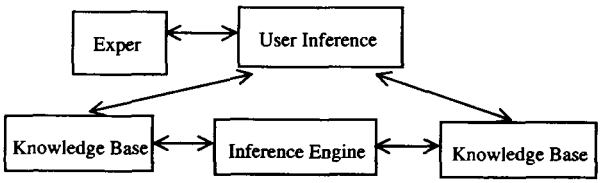

图 1 数据库推理问题控制器原型

为了在数据库会话(session)中阻塞推理通道，对用户的查询可以采取的方法主要有：在查询执行前进行适当的修改；对查询结果进行修改等二种。对查询结果进行修正是当前研究的重点，因为恶意用户可以通过一系列未受限制的查询，推理出不允许查询的信息。

图 1 是一种典型的数据库推理问题控制器的原理示意图。该控制器由系统安全员使用的基于知识的工具，来检测和制止推理。推理引擎根据规则进行以下动作：(a)判断哪些查询及相关结果返回给用户；(b)检查返回给用户的数据的完整性和一致性；(c)启动对推理风险的概率的计算并确认推理通道；(d)根据从用户来的数据和(c)的结果修改知识库；(e)对安全级别的修改提出建议。

4.2 多实例

多实例是指在多级安全数据库管理系统中同时存在具有相同名字的多个元组(元组级多实例)或多个数据元素(元素级多实例)，这些元组或元素的区别在于安全级别。一个具有多实例元组的关系包含具有相同的外在主关键字和不同安全级别的多个元组。一个元素级多实例的关系包含两个或更多的元组，这些元组具有相同的外在主关键字和安全级别值，但是其它元素不同。

多实例是多级安全数据库系统为了实现 MAC 机制而采取的一种技术，它对源于关键字完整性的推理问题有效。但多实例导致了数据库的信息保密与数据库的完整性间的矛盾。用户可能面对多个视图，而这些视图包含了矛盾的数据；另外，一个具有多实例特性的数据库系统需要一定的机制和管理程序周期性地清除不需要的多实例元组，这对用户造成了额外的负担，也增加了数据库管理的难度。

4.3 审计

DBMS 提供对用户行为足够强的审计能力以检测推理问题。在某些系统中，审计被用来控制推理问题，例如可以保存一个用户的查询历史记录，每当用户进行查询，就对历史记录进行分析以判断当用户的查询响应与以前的查询相关联时是否产生推理问题。

用审计方法检测推理问题存在以下缺点，由于基于这样的假设：通过分析审计记录来发现用户的异常行为，检测推理通道，因此检测出的推理通道比较有限；长时间的保持历史文件必然影响到所能记录的审计数据。

5 数据级的推理控制方法

在安全数据库系统中，少数推理是确定的，如函数依赖；然而在多数情况下，推理是不确定的，或者说是具有一定概率

(下转第 198 页)

下列的图像(如图 11)。

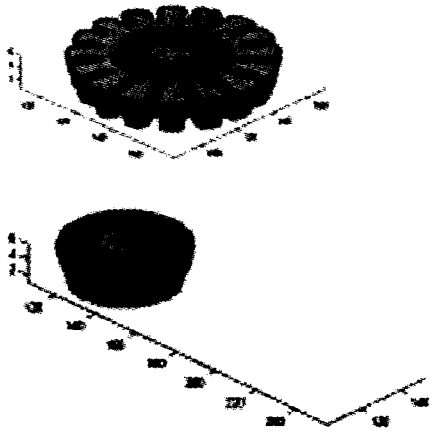

图 11 将三维重建的图像按层剥离效果图

结束语 工业 CT 在无损状态下能获得被检断面的二维灰度图像,从图像上可以直观地看到目标细节的位置、形状、大小。本文提出的三种断层图像的边缘检测方法,具有不同的适用范围,可以根据具体要求选择。断层图像的边缘检测不仅在医学领域具有重要价值,在工业无损探伤天体结构的观察,及其他更广泛的领域也有着极其重要的实用价值和研究意义。

参 考 文 献

1 Levoy M. Display of surface from volume rendering. IEEE Computer Graphics and Application, 1998,8(5):2937

2 Fessler J A. Statistical image reconstruction methods for transmission tomography. In:SPIE Handbook of Medical Imaging, M. Sonka, J. M. Fitzpatric, eds. SPIE Press. Bellingham, Washington, USA, volume 3,2000

3 Leahy R, Editorial B C. Recent development in iterative reconstruction for PET and SPECT. IEEE Trans. Medical Imaging, 2000,19:257～260

4 Fessler J A, Ficaro P E, Clinthorne N H, Lange K. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. IEEE Trans. Medical Imaging, 1997, 16:166～175

5 Liu W C, et al. Dynamic elastic interpolation for 3D medical image reconstruction from serial cross-sections. IEEE Trans. Med. Imaging,1988,7(3):225～232

6 Liu W C, et al. A new surface interpolation technique for reconstructing 3D objects from serial cross-sections. Computer Vision, Graphics, Image Processing,1998,48(4):124～143

7 Sofeman Z, et al. Advanced graphics behind medical virtual reality: evolution of algorithms, hardware and software interface. Proceeding of IEEE, 1998,86(3):531～554

8 彭群生,等著. 计算机真实感图形的算法基础. 北京科学出版社, 1999

9 李燕,谭鸥,段会龙. 三维医学图象可视化技术综述. 中国图象图形学报,2001,6(2):108～109

10 王新成. 高级图像处理技术. 中国科学技术出版社,2001

11 孙兆林. MATLAB 6. X 图像处理. 清华大学出版社,2002

(上接第 112 页)

的。数据级的推理控制方法的思想是:通过数据库中的数据自身来反映数据库中的推理概率。以函数依赖为例,分析数据库中的数据,生成函数依赖集合用于推理检测。

在数据级的推理控制方法中,利用数据挖掘方法[8](数据挖掘就是从大量的、不完全的、有噪声的、模糊的、随机的数据中,提取隐含在其中、人们事先不知道的、但又是潜在有用的信息和知识的过程)来发现数据库系统中的推理规则是一个非常有前途的研究方向。数据挖掘技术还可以被用来增强数据库系统的安全性,如对审计记录进行分析,攻击检测等[9]。

但数据挖掘技术对数据库的安全带来了新的威胁:数据挖掘生成的规则可能引起新的推理通道。因此,如何在数据挖掘中防止非法的信息泄露也是值得研究的一个问题。

总结 本文介绍了多级安全数据库系统中存在的推理问题,用于推理的信息来源及方法策略;研究并分析了目前的研究方向和内容,主要包括:推理的形式化,数据库设计中的推理控制技术,数据库运行时的推理控制方法,以及数据级的推理控制方法,并指出了各个技术的优缺点。

参 考 文 献

1 Hinke T H, Harry S. Protecting databases from inference attacks [J]. Computer & Security, 2000,16(8): 687～708

2 Li Lixin, Liao Changrong, Chen Weiming, et al. Inference attack analysis of MLS DBMS using rough set theory[A]. In: Proc. of 7th Joint Intl. Computer Conf. Shantou, China, 2001. 1339～1452

3 Wang L, Wijesekera D, Sushil J. Cardinality-based inference control in sum-only data cubes[A]. In: Proc. of the 7th European Symposium on Research in Computer Security, 2002

4 Domingo-Ferrer J. Advances in inference control in statistical databases: An overview in inference control in statistical databases: from theory to practice[A]. LNCS 2316, Springer-Verlag, 2002. 1～7

5 Su T, Ozsoyogiu G. Data dependencies and inference control in multilevel relational database systems[A]. In: Proc. of the IEEE Symposium on Security and Privacy, 1987. 202～211

6 Millen J. Honesty is the best policy[A]. In: Proc. of the 3rd Rome Laboratory Database Security Workshop, 2001

7 Rizvi S, Haritsa J. Maintaining data privacy in association rule mining. In: Proc. of the 28th Intl. Conf. on Very Large Data Bases, 2002. 682～693

8 Evfimievski A, Gehrke J, Srikant R. Limiting privacy breaches in privacy preserving data mining. In : Proc. of the 22nd Symposium on Principles of Database Systems, ACM Press, 2003. 211～222