元数据相关推理研究*)

毛奇正　柏文阳　刘奇志

(软件新技术国家重点实验室　南京210093)　(南京大学计算机科学与技术系　南京210093)

摘　要　在多级安全数据库中，推理通道的存在会对信息的安全造成威胁。在推理问题中，与元数据相关的推理是其中的一个重要方面，对该问题的研究有助于数据库的安全增强。将与元数据相关的推理问题进一步划分为不同类别描述，对不同类别的推理问题，分别探讨了如何进行控制，并给出了解决的方法。为了消除推理通道，需要修改属性的安全级别标识信息；提出了一个标识修改的模型MTL，用于通过修改属性安全标识来消除推理通道。

关键词　数据库安全，推理控制，函数依赖，元数据

Research on Inference Problems Based on Metadata

MAO Qi-Zheng　BAI Wen-Yang　LIU Qi-Zhi

(State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093)

(Department of Computer Science and Technology, Nanjing University, Nanjing 210093)

Abstract　Database security is a topic studied by many people. When dealing with this topic, the affect of inference channels is a problem that should not be avoidable. Inference channels bring menace to the information in database, and the widely used multilevel model cannot ensure the security of the whole system. Many of the inference problems are based on metadata of the database system. Those problems are partitioned into different classes and discussed according to their own characteristic in this paper. To eliminate inference channels, the security level of the attributes must be modified. A model (MTL) which could be used in this procedure is given.

Keywords　Database security, Inference channel, Metadata, Functional dependency

1　引言

在数据库系统中，由于数据的集中管理以及随之而来的多用户存取和近年来迅速发展的跨网络的数据库系统，使得数据库的安全问题变得极为突出。数据库安全也自然成为信息安全的一个研究重点。现在为了保证数据库的安全，安全数据库往往都使用多级安全(Multilevel Secure)模型来对数据库进行存取控制，从而构成一个多级安全数据库系统。

强制访问控制策略使数据库具有多级安全性(MLS)，但是强制访问控制并不能彻底保证信息的安全。可以看到，因为数据之间存在着各种语义联系，即使在一个强制访问控制完全发挥效用的系统中，也会因推理通道的存在而威胁到信息的安全。控制推理通道成为数据库安全研究不得不面对的难题。因此，必须对多级安全数据库中所存在的推理通道进行控制，以避免上述因推理而产生的信息泄露现象[1,2]。

数据库中的推理需要使用外部知识。元数据作为数据的数据，包含了数据库中数据的描述信息，是使数据发挥作用的重要条件之一，可以被用来发现推理通道，造成数据库中的信息泄漏，因此要对此类推理问题进行控制。

2　多级安全数据库中的推理问题

2.1　多级安全数据库

多级安全数据库通过使用基于多级安全模型(如 Bll-Lapadula 模型，简称 BLP 模型)的强制访问控制(Mandatory Access Control，简称 MAC)机制来保证数据库的安全。多级安全模型所使用的安全策略是基于被分配给数据库中的主体和客体的安全分类标识的，所有的安全分类标识构成了一个安全格(Lattice)。采用基于多级安全模型的强制访问控制策略有利于保证信息流按照由低到高的方向流动，加强了对数据库的存取控制。

2.2　推理问题

在多级安全数据库中，即使强制访问机制完全发生作用，由于推理通道的存在，有时也不可避免地存在信息泄漏。在多级安全数据库中推理是指，数据库中拥有低安全分类标识的用户，通过获取低安全分类标识的数据可以推断出具有高安全分类标识的信息[3]。Marks[4]给出了一个数据库推理问题的正式的定义：

多级安全数据库中的推理是指用户从数据库中获取元组集 T，该元组具有属性 A，此时如果可以描述出一个元组集 T'，具有属性集 A'，而且有 T' 不包含于 T 或者 A' 不等于 A。从逻辑上讲，此时存在一个推理通道，从元组集 T 可以推导出元组集 T'。

因为低权限的用户只是访问属于自己权限范围的数据，此时不违反强制访问控制策略，但如果此时元组集 T' 中又包含了对于该用户为具有高安全分类标识的信息，此时该用户就可以获取高于自己权限许可的信息，不可避免地造成信息泄漏。这便是由于推理通道的存在，使得高安全分类标识的信息可以通过推理得到，从而造成信息泄漏。

3　与元数据有关的推理问题

3.1　元数据的内涵

元数据最简短的定义是“数据的数据”。许多专家和学者

*)国家“八六三”高技术研究发展计划基金项目(编号2002AA141091)。毛奇正　硕士，主要研究领域为数据库安全、数据挖掘、数据仓库。柏文阳　副教授，研究方向：数据库安全、数据挖掘、数据仓库。刘奇正　博士，讲师，主要研究方向为网络环境下的数据管理与工业测控领域的数据管理技术等。

从不同的侧重点出发给元数据以不同的描述。例如，认为元数据是数据库管理领域的概念，是关于数据组织的数据；或者认为元数据是对数据的描述以及对数据集当中的数据项的解释，它能提高数据的利用价值。

数据库中的元数据中储存关于数据的内容、质量、状况和其他特性的信息，包括数据模式的信息，数据库中表、视图等客体的描述信息，数据库的所有者，完整性约束等。元数据是使数据发挥作用的重要条件之一，它帮助数据生产单位有效地管理和维护数据。

3.2 和元数据相关的推理

多级安全数据库中的推理往往都是由于查询获得的结果数据和用于存储、管理这些数据的元数据综合作用的结果。这里的元数据主要是指数据库的完整性约束。

完整性约束提供了一种手段来保证用户对数据库作修改时不会破坏数据的一致性。一般包括：一个或多个属性取值的值域的约束；主关键字的完整性；外关键字引用完整性；断言；触发器以及函数依赖关系。

通过元数据进行推理的推理通道大致可以分为以下三种。

3.2.1 *推测客体的存在* 这主要是通过关键字完整性约束来达到。关键字完整性保证了在关系中每个元组的唯一性，从而减少了数据冗余。当关系中的所有关键字具有相同的安全分类标识，此时约束不会产生推理相关的安全问题。但是当一个低安全分类标识的用户希望增加关系中的一个元组时，此时关系中数据被标识为元组或者字段一级的安全。如果此时具有同样关键字，并且具有高的安全分类标识的元组已经存在，为了保证数据库关键字唯一性约束，数据库必须删除已经存在的元组或者通知用户具有相同关键字的元组已经存在。但无论是在前面的哪一种情况下，都会出现问题：在第一种情况下，低安全分类标识用户的新增操作就会导致高安全分类标识用户插入的数据丢失，是不可接受的(这样不会产生相应的信息泄漏，但是他可能导致严重的完整性问题，或者导致拒绝服务的攻击)；而在第二种情况下，就出现了推理通道，因为高安全分类标识数据的存在影响低安全分类标识用户数据的存在性。

外关键字引用完整性引起的推理与此类似，可以推测外关键字所引用的哪个属性中某个值存在与否。

3.2.2 *推测数据库中的规则* 在实际应用中，数据库中的数据往往要满足一些条件，例如雇员的年龄应在18～60之间。元数据中的断言、触发器、属性值的值域等元数据用于记录这些限制条件。安全数据库不仅要对存储的信息进行保护，也要保护这些限制规则。

当一个用户对数据库中的内容进行增加、删除、修改操作的时候，如果违反了这些限制，操作将不被接受，但用户可以利用这个来对规则进行推理。例如，在某个雇员表中，“工资”属性的上限被定为3000，用户通过不断地插入元组可以发现，“工资”属性的值超过3000的都不被接受，便可以猜测存在这么一条限制规则。

3.2.3 *推测数据库中数据的取值* 在实际应用中，这是对数据库中的信息安全威胁最大的一类推理问题。推测数据的取值可以通过函数依赖约束来进行，也可以通过涉及到多个数据项上的值的约束关系来进行。

Su 和 Ozsoyoglu[5~7]在文章中首先提出函数依赖问题。作者研究认为推理通道出现的原因更多是因为函数依赖和多值依赖这些在一个关系属性上的约束。下面是一个典型的函数依赖的例子：

假设存在定义为(NAME，RANK，SALARY，EXEPERLENCE)的一个关系模式，并且假设关系 Employee 基于这个模式，而且存在函数依赖 RANK→SALARY。此时的安全要求只有 Top-secret 安全分类标识的用户才能了解一个雇员的薪水情况，也就是说如果仅仅拥有 Secret 或更低的安全分类标识的用户不能同时取得属性 NAME 和 SALARY 的值，即在属性集〈NAME，SALARY〉上使用关联约束。但同时也要保证这个分类模式允许具有 Secret 或更低安全分类标识的用户可以单独选择这样的属性字段，以便支持数据的可用性。此时可以发现，在对于直接存取数据库的安全得到保证以后，而通过间接存取仍然可以危害数据的安全。

涉及到多个数据项上的值的约束关系。如果一个约束所涉及的数据项具有不同的安全分类标识，约束的使用就会导致推理通道的存在，下面是一个例子[8]：设属性 A 的安全分类标识为 Unclassified，而属性 B 的安全分类标识为 Secret。此时有约束保证 A＋B≤20，对于 Unclassified 安全分类标识的用户是可用。B 的值并没有直接影响 A 的值，但是由于约束关系的存在，决定了 A 所取的可能值。此时就存在了推理通道。

4 元数据相关推理通道的控制

上述的三种推理问题中，对规则的推理一般不会单独对数据库中的信息造成实质性的威胁，对此进行控制势必会影响数据库的可用性。故本文只针对其余的两种推理问题进行研究。

4.1 对客体存在性的推理控制

对于这个问题，一般采用多实例的方法来解决。多实例(Polyinstantiation)是指同时存在多个同名但是安全级别不同的数据客体。这些实例按照它们的安全级来区分。多实例是只有多级数据库中才有的现象，在标准的关系数据库中，由于主关键字的唯一性，因此不会出现重复的数据客体。但是在多级安全数据库中，必须允许关键字相同的元组的出现，否则如果低级别的用户试图插入新的数据，而此数据在关系中已经存在，并有较高的存取级，这时如果拒绝插入，用户可以据此判断出在关系中有高安全级的数据客体存在，因此必须允许插入，而又不能更改老的数据，因此不同存取级的数据必须同时存在。具体的处理如下：

Ⅰ)当主体向关系中插入一个元组，而关系中已经存在一个具有相同主关键字的不可见元组时，产生一个多实例元组。

Ⅱ)当主体更新某元组而该元组对主体是部分可见时，即元组中的部分属性的存取级比主体存取级高时，产生一个多实例元素。

4.2 对具体数值的推理控制

对于涉及到多个数据项上的值的约束关系所引起的推理问题，可以调整各个数据项的安全标识到同一级别，如果不能通过调整数据项的安全级别来消除，通常的解决方法是只允许该约束定义在单个的安全分类标识上。需要把定义在几个具有不同的安全分类标识的数据项上的约束，分解成几个单独数据项上的约束。例如约束 A＋B≤20，可以分解成 A≤10 和 B≤10。还可以在数据库的查询阶段来对用户的查询进行动态检查，如果用户查询会造成信息泄漏，则拒绝该查询。

为了消除因为函数依赖引起的推理问题，应该先通过函

数依赖关系和所涉及到的属性的安全级别，来发现推理通道。如果存在函数依赖 A→B，当 A 的安全级别低于 B 的安全级别时，才认为存在推理通道；但是在关系数据库中，关键字常被用来作为查找的依据，因此关键字的安全级别一般定义得较低，而在函数依赖 A→B 中，如果 A 是关键字，则即使 A 的安全级别低于 B 的安全级别，由于 Unique 关键字（包括主关键字）的唯一性约束要求一个 Unique 关键字只能对应一个元组，不会产生推理通道。

发现推理通道后，要给出如何修改属性的安全标识的建议。具体作法是提升决定因素的安全标识使得至少有一个的安全标识使之不小于被决定因素。但是安全分类标识的修改存在多种选择，而且修改安全分类标识，必然降低部分信息的可用性，同时还要注意提升某个属性的安全标识时，可能会引起新的推理通道或者一并消除现有的推理通道。理想的标记修改的方法，应该是可用信息丢失最少，一次修改消除尽可能多的推理通道，同时新引起的推理通道最少。

为了满足上述要求，依据图论的思想，设计安全标识修改的模型 MTL。在此模型中，每个属性用一个点表示；对每一条函数依赖关系，从决定因素中的每一个属性向被决定因素引一条有向边，其权值为决定因素中元素个数的倒数。例如，对于函数依赖关系 A，C→ B，从点 A 和 C 各引一条边到 B，每条边的权值为1/2。

为每个属性定义“入度(in)”和“出度(out)”。“入度”定义为在该表上的函数依赖关系中作为被决定因素的次数。计算方法是将指向该店的各边权值相加；“出度”定义为作为决定因素的次数，其值为从该点出发的边的数目。

对于安全分类标识的提升，比如从1级提升到2级，和从3级提升到5级所造成的信息丢失是不同的，因为低级别的属性可能为更多的用户所共享，而提升该安全分类标识可能造成更多的信息可用性丢失。对于函数依赖关系左边的每个属性，定义“跨度(span)”为该属性安全标识与右边属性的安全标识的差，即对于推理规则 A →B，A 中的属性 x，span(x)＝level(B)-level(x)。

以上的几个因素需要综合考虑，“入度”越小，则可能引起新的推理规则的几率越小；“出度”越大，则提升该属性的标识越可能连带消除其他的规则；跨度越小，则提升该属性的标识所引起的可用性丢失越小。可以给它们赋予不同的权重 w1，w2和 w3，然后定义一个属性的权为 weight(x)＝ (w1 * span (x)＋w2 * in(x))/(w3 * out(x))，提升权值最小的属性的安全标识。

算法的描述：对推理控制检查得到的规则集中的每一条规则，计算规则左边（决定因素）各属性的权值，选择权值最小的属性提升其安全标识，如果该属性的入度不为0，则对于该属性作为被决定因素的函数依赖关系，检查是否会生成新的推理规则，如果会，则将该规则加入规则集。

进行推理控制检查的算法如下：

```
errorset={}; Pkset={}; fdset={};
GetPK(schemaname,tablename,PKset);
GetFD(schemaname,tablename,fdset);
for each fd in fdset {
  for each x in fd.forepart {
    if x not in Pkset {
      if x.level>m then m=x.level; } }
  if m<fd.hindpart.level then
    errorset=errorset∪fd
}
```

标识修改的算法如下：

```
for each e in errorset {
   x=minWeight(e.forepart);
   x.level=e.hindpart.level;
   if x.in!=0 check(related(x));
}
```

其中 minweight(A)用于查找属性集 A 中权值最小的属性，related(x)找出那些以属性 x 作为规则右边（被决定因素）的函数依赖关系。

在给出修改标识的建议后，如果确是因实际需要而不能修改，则可以将该规则储存起来留在运行期进行推理检查时使用。

结束语 本文为在数据库的设计期进行与元数据相关的推理控制提供了一种较为完整的解决方案，为运行期的推理控制工作奠定了基础，从而为实现更高安全级别的安全数据库系统中的推理控制提供了可能。该方案目前被应用于我们所承担的国家八六三计划信息安全领域内的数据库安全增强技术课题中，以期实现一个能够提供 B2级（我国国家标准的结构化保护级）安全标准的主要安全功能的安全增强器，取得了很好的效果。当前，由于很难找到一个完整的理论体系来描述所有的推理问题，元数据的种类又名目繁多，因此还需要更多的研究工作，而且为了达到推理控制的目的，还要在运行期进行推理检查。

参 考 文 献

1 Denning D E,et al. Views for Multilevel Database Security. IEEE Transactions on Software Engineering,1987,13(2): 129～140

2 Denning D E. A Preliminary Note on the Inference Problem in Multilevel Database Management Systems. In: Proc. of the National Computer Security Center Invitational Workshop on Database Security, June 1986

3 Stickel X,et al. Detection and Elimination of Inference Channels in Multilevel Relational Database Systems. In: Proc. 1993 IEEE Symposium on Security and Privacy,1993

4 D G. Inference in MLS Database Systems. IEEE Trans. on Knowledge and Data Engineering,1996,8(1)

5 Su T. Inferences in Database: [Ph. D. Dissertation]. Department of Computer

6 Su T,Ozsoyogiu G. Data Dependencies and Inference Control in Multilevel Relational Database Systems. In: Proc. of the IEEE Symposium onSecurity and Privacy,1987. 202～211

7 Su T,Ozsoyoglu G. Multivalued Dependency Inferences in Multilevel Relational Database Systems. In: Database Security III: Status and Prospects, eds. D. L. Spooner and C. Landwehr, pp. NorthHolland, Amsterdam,1990. 293～300

8 Meadows C, Jajodia S. Integrity Versus Security in Multi-Level Secure Databases. In Database Security: Status and Prospects, Carl E. Landwehr,ed North-Holland, Amsterdam,1988. 80～101

9 Engineering and Science, Case Western Reserve University, Aug. 1986