基于关系数据库的极松散结构数据模型的访问控制研究

潘 颖[1,2],汤 庸[3],刘 海[3]
(1. 中山大学信息科学与技术学院,广东广州 510006;2. 广西师范学院计算机与信息工程学院,广西南宁 530023;
3. 华南师范大学计算机学院,广东广州 510631)

摘 要: 本文提出一个针对数据空间环境下极松散结构模型的细粒度和动态的访问控制框架:首先定义通用的极松散结构模型GLSDM(General very Loosely-Structured Data Model);给出GLSDM到关系表的映射方法,将GLSDM上细粒度的访问控制转换为关系表的row、cell等安全级别的访问;通过动态查询重写,在用户执行查询时将GLSDM的权限信息添加到SQL语句中,从而实现GLSDM的动态访问控制.理论和实验证明该框架是可行和有效的,本文的映射方法和动态查询重写算法能够保证对GLSDM的访问控制通过等价的关系数据库的访问控制来实现.

关键词: 访问控制; 数据空间; 关系数据库; 松散结构
中图分类号: TP309 **文献标识码:** A **文章编号:** 0372-2112 (2012) 03-0600-07
电子学报URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2012.03.032

Access Control in Very Loosely Structured Data Model Using Relational Databases

PAN Ying[1,2], TANG Yong[3], LIU Hai[3]
(1. *Department of Computer Science, Sun Yat-Sen University, Guangzhou, Guangdong* 510006, *China*;
2. *College of Computer and Information Engineering, Guangxi Teachers Education University, Nanning, Guangxi* 530023, *China*;
3. *School of Computer Science, South China Normal University, Guangzhou, Guangdong* 510631, *China*)

Abstract: This paper proposes a framework to efficiently support dynamic, fine-grained access control for the very loosely structured data model, named GLSDM(General very Loosely-Structured Data Model), which is presented based on the current dataspace data models. In the framework, GLSDM is mapped into and stored in relational databases, and then the fine-grained access control in GLSDM is converted into the corresponding fine-grained security (e.g., row-level and cell-level security) in relational databases. A query rewriting algorithm is also given to dynamically imbed GLSDM security information into SQL statements the user issues, thus, dynamic access control is realized during the period of query processing. Finally, the validity of the framework is proved by theory and experiment, that is, the GLSDM-to-relational mapping method and query rewriting algorithm in this paper can ensure the access control in GLSDM is equivalent to that in relational databases.

Key words: access control; dataspace; relational databases; loosely structured

1 引言

2005年Franklin等人提出数据空间(dataspace)概念[1],引起人们对极松散结构数据模型(very loosely structured data model)的关注.数据空间是一种新型的数据管理模式,主要思想是投入极低的前期成本将与某个组织或个体有关的一切信息集成起来,对这些信息实现pay-as-you-go的管理,即:系统开始提供的是关键字查询之类的简单功能,当用户需要时,系统才会投入更多的成本做一些必须的工作(如集成更多信息、提高数据间语义映射的质量等),进而提供更强的服务[2].和传统的集成系统和RDBMS相比,数据空间更适合管理现实世界分布、异构的和动态变化的数据.数据空间中的极松散结构的数据模型没有严格、统一的schema,相对其他数据模型(如XML和关系模型),它能够更容易且花费更少的代价来描述各种数据,更有利于对数据实现pay-as-you-go的管理.

数据空间的访问控制有特别的要求:(1)数据空间需要描述各种粒度的数据源,因此访问控制必须是细粒度的(fine-grained),能够对不同层次、不同粒度的数据

收稿日期:2010-12-14;修回日期:2011-10-27
基金项目:国家自然科学基金(No.60970044, No.60673135, No.60736020);广东省自然科学基金(No.7003721);广西教育厅科研项目(No.200807LX429)

进行访问.(2)数据空间对数据源的描述是动态和渐进的过程,访问者的权限也随着数据属性、环境条件等因素动态变化.因此,数据空间必须支持动态访问.(3)数据模型的极松散性意味着被访问的对象没有固定的模式,同时模型对数据及其关系的描述会存在不完全信息,权限访问往往涉及到数据及其复杂关系,数据空间的访问控制必须能应对这些由模型的松散性带来的问题.

由于极松散结构数据模型和以往的数据模型不同,现有的访问控制技术不能直接用到该模型中.本文针对数据空间访问控制的需求和极松散模型的特点,提出一个细粒度和动态的访问控制框架:定义一个通用的极松散结构数据模型 GLSDM(General very Loosely-Structured Data Model);给出基于关系数据库(RDB)的访问控制方法,通过对关系表的等价的访问控制来实现对 GLSDM 的细粒度和动态访问.最后给出理论证明和实验分析表明本文的方法是可行和有效的.

2　相关工作分析

2.1　数据空间的访问控制

目前针对数据空间的访问控制的研究很少.文献[3]对数据空间模型 iDM(iMeMex Data Model)[4]进行了扩展,添加 access component 和 authorization component 来描述权限信息,从而支持用户对 iDM 的安全访问.和该文不同,本文不在数据模型上标注权限信息,而是利用 RDB 的访问技术来实现访问控制.此外,本文的方法不是针对特定的模型 iDM,而是面向通用的极松散结构数据模型,因而会更具普遍性.

2.2　基于 RDB 的 XML 文档的访问控制

文献[5]使用基于路径的方法将 XML 文档映射并保存到关系表中,然后利用 RDB 的访问控制技术对 XML 文档进行访问控制;该作者还在文献[6]中给出 RDB 正确支持 XML 访问控制的基本条件:关系模型和 XML 模型具有等价的访问控制对象及等价的 deep set 操作.文献[7]研究基于 Native XML 和关系数据库的 XML 文档的访问控制,并提出使用重新注释的方法来计算 XPath 查询.然而,由于数据空间要求更灵活和更复杂的细粒度和动态的访问,以及 XML 和 GLSDM 两者的不同(如:XML 不够松散且节点间的关系较简单;XML 是树状模型而 GLSDM 是图模型等),因此基于 RDB 的 XML 文档的访问控制技术不能直接用到 GLSDM 中.本文主要做了如下和以往研究不同的工作:给出 GLSDM 到关系表的映射方法,该方法适合 GLSDM 极松散的特点并考虑了不完全信息的映射;实现基于 RDB 的 GLSDM 的细粒度和动态访问控制(以往文献都没有涉及基于 RDB 的 XML 文档的动态访问控制的具体实现问题);为了更好的和 RDB 的查询技术结合,本文没有采用 deep set 操作而是采用了 SQL 来描述访问控制对象和访问控制规则,而且本文的安全性分析和证明方法也和文献[6]不同.

2.3　RDB 的细粒度和动态访问控制

文献[8]研究了在 Hippocratic 数据库中实现 cell 级别的访问控制的方法;文献[9]通过在 grant 语句添加 predicate 断言来实现 cell 级别的访问控制.这些文献都没有具体讨论动态访问控制问题,同时实现 RDB 的细粒度和动态访问控制的研究有:文献[10]提出了一种能用 SQL 描述和执行的动态的细粒度访问控制策略;Oracle 的 VPD(Virtual Private Database)技术[11]也提供了动态的细粒度访问控制.此外,有些文献研究细粒度角色委托模型,以保障数据库系统安全[12,13].我们可以借鉴上述工作来实现 GLSDM 的细粒度和动态访问控制,但由于关系模型和 GLSDM 两者的结构不同,如何有效地将 GLSDM 的访问控制对象映射到关系模型的访问控制对象是一个需要解决的问题.

3　通用的极松散结构数据模型 GLSDM

数据空间模型 iDM[4]通过元组 γ_i 描述数据间的序列关系(sequence relationship),尽管复杂关系可以看作序列关系的组合,但用 γ_i 描述复杂关系不够简单明了.文献[14]的模型是类似 RDF 的带标签的图,由于 RDF 不够松散,这种模型严格来说不是极松散模型.文献[15]提出了由实体的 attribute-value(属性-值)组成的模型,它将边(关系)也作为 attribute-value 处理,这样会使模型更简单,但同时也混淆了实体自身具有的属性和实体间关系的不同.文献[16]提出了 iDM 的简化模型,但对不完全信息的描述考虑不够.在这些模型的基础上,我们提出一种更为通用的数据空间模型,它的形式定义如下:

定义 1　GLSDM 是用来描述包含在数据空间中的数据及其关系的图模型,记为 $G:=(N,E)$,其中 N 为节点集$\{N_1,\cdots,N_k\}$,节点 N_i 由 attribute-value 组成,记为 $N_i=\{(A_1^i,V_1^i),\cdots(A_n^i,V_n^i)\}$,当 $N_i=\emptyset$ 时,称 N_i 为空节点.E 是边的集合,边记为(N_i,N_j,L),其中 $N_i,N_j\in N$,$i\neq j$,L 表示边的名称,且 L 可为 null 值.

GLSDM 有如下特点:(1)GLSDM 能够描述数据源中各种粒度的逻辑实体及其关系.例如,节点可描述整个文档,也可以描述文档中的章节、小段文字等不同粒度的逻辑实体;边描述这些逻辑实体之间的各种关系.(2)GLSDM 是结构非常松散的模型,具体表现在数据没有统一的 schema:(a)不要求节点的属性相同,如 N_1、N_2 均描述"学者"这类实体,它们的属性列表可以不同.

甚至可存在没有属性的节点,即空节点.空节点可用来描述不完全信息,如已知存在某一节点,但节点的具体信息未知时,可用空节点表示.(b)节点间的联系是随意的,用 L 可表示不同的关系.当关系存在但具体是何种关系不确定时,L 为 null 值.当节点间都不存在关系时,$E=\emptyset$.和其他节点没有任何关系的空节点是没有意义的,因此,本文中的空节点总有边和其他节点联系.

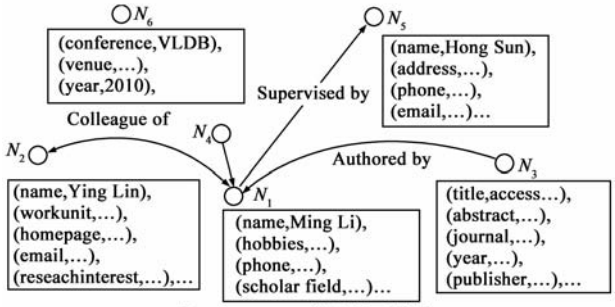

图1 GLSDM描述学术信息

用 GLSDM 描述部分学术信息的例子如图 1 所示,其中,N_1、N_2、N_5 描述学者的信息,N_3、N_6 分别描述论文和会议的信息,N_4 是空节点.

4 基于 RDB 的 GLSDM 访问控制框架

基于 RDB 的 GLSDM 的访问控制框架如图 2 所示:(1)首先将 GLSDM 映射到关系表;(2)对 GLSDM 的细粒度访问控制转换为对相应关系表的访问控制,即对 GLSDM 不同粒度的对象的访问转换为对相应关系表的 row、cell 等安全级别的访问;(3)将 GLSDM 的访问规则映射到 RDB 的访问规则,通过动态查询重写,在用户执行查询时将 GLSDM 的权限信息添加到用户的 SQL 语句中,从而动态控制用户对数据的访问;(4)将关系表的查询结果转换成 GLSDM 上的查询结果.

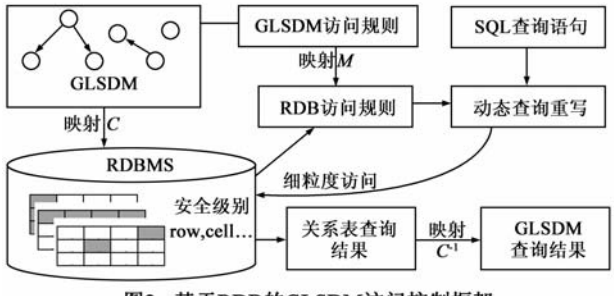

图2 基于RDB的GLSDM访问控制框架

4.1 GLSDM 到 RDB 的映射

定义 2 GLSDM 中的数据记为 D_G,关系表中的数据记为 D_R,则 GLSDM 到 RDB 的映射记为 $C(D_G)=D_R$.由 GLSDM 的定义可知,D_G 由 GLSDM 边的数据 E 和节点的数据 attribute-value 组成,映射 C 将它们分别映射到不同的关系表中,具体规则如下:

边的映射规则:将 GLSDM 中的边映射到表 Edge 中,该表的关系模式为:Edge (source, label, target).其中,source 和 target 字段分别用来存储边的引出节点和引入节点,label 字段用来存储边的名称.

节点 attribute-value 的映射规则:将 GLSDM 节点的 attribute-value 映射到表 Attribute(nodeID, value).其中,nodeID 字段存储节点的 ID,value 字段存储节点的属性值.我们将属性名作为表名,为每一种属性单独设计一个值表.为描述简便,本文统称这类表为表 Attribute.

映射时涉及的不完全信息的情形有:(1)边的名称不能确定时,将边映射到表 Edge 的元组的同时,设 label 字段的值为 null.例如图 1 的边(N_4, N_1, null),映射的结果为 Edge(N_4, null, N_1).特别的,当 GLSDM 所有节点间都没有关系时,映射结果只有表 Attribute,没有表 Edge.(2)空节点映射时,只需将与其相连的边映射到表 Edge 即可.特别的,当 GLSDM 所有的节点都是空节点时,映射结果只有表 Edge,没有表 Attribute.

Email Name ...

source	label	target
1	Supervisedby	5
4	null	1
3	Authoredby	1
...	...	...

表 Edge

nodeID	value
2	yinglin@...
5	hongsun@...
...	...

nodeID	value
1	Ming Li
2	Ying Lin
5	Hong Sun
...	...

表 Attribute

图 3 GLSDM 映射到表 Edge 和表 Attribute

按以上方法可将图 1 的 GLSDM 映射到图 3 的表 Edge 和表 Attribute.

4.2 细粒度访问控制的实现

RDB 的细粒度访问控制是指访问控制可限制到数据表的行(rows)、列(columns),甚至元素(cells).在 GLSDM 中,细粒度访问控制要能限制到 GLSDM 的边和节点的 attribute-value.当 GLSDM 映射到关系表后,GLSDM 的访问控制要通过 RDB 的访问控制来实现.

定义 3 访问控制对象表示被访问的资源,记为 object.

定义 4 关系模型的访问控制对象记为 object_R.为方便用 SQL 描述 object_R,我们定义 object_R 为一个 3 元组(F, T, P),分别对应 SQL 的 SELECT、FROM 和 WHERE 子句.其中,T 代表权限涉及到的表(table)的集合;F 代表权限约束到的字段(fields)的集合,且 F 包含在 T 的字段集合中,记为:$F \subseteq T.\text{fields}$;$P$ 是与 T.fields 有关的谓语(predicate),代表限制条件,当无限制条件时,P = null.

定义 5 object_R 的 SQL 表示,记为 $\text{object}_{\text{R-SQL}}$ = select F from T where P.

例,$T = \{T_1, T_2\}$, $\text{object}_{\text{R-SQL}}$ = select F_{T1}, F_{T2} from T_1, T_2 where $P_{T1} \wedge P_{T2} \wedge P_{\text{join}}$,其中 P_{join}是表间的连接谓词.

定义 6　GLSDM 的访问控制对象记为 object_G,根据 GLSDM 的结构特点,object_G 可细分为:$\text{object}_{\text{G-all}}$、$\text{object}_{\text{G-edge}}$、$\text{object}_{\text{G-AV}}$和 $\text{object}_{\text{G-edge\&AV}}$,分别代表权限约束的粒度为:整个 GLSDM、整条边或边中的某部分、节点的全部或部分 attribute-value、整条边或部分边以及节点的全部或部分 attribute-value.

定义 7　object_G 到 object_R 的映射,记为 $M(\text{object}_G) = \text{object}_R$.定义映射 M 如下:

$$M(\text{object}_{\text{G-all}}) = (\text{edge.fields} \cup \text{attribute.fields}, (\text{edge}, \text{attribute}), \text{null}) \quad (1)$$

$$M(\text{object}_{\text{G-edge}}) = (F \subseteq \text{edge.fields}, \text{edge}, P_{\text{edge}}/\text{null}) \quad (2)$$

$$M(\text{object}_{\text{G-AV}}) = (F \subseteq \text{attribute.fields}, \text{attribute}, P_{\text{attribute}}/\text{null}) \quad (3)$$

$$M(\text{object}_{\text{G-edge\&AV}}) = (F \subset \text{edge.fields} \cup \text{attribute.fields}, (\text{edge}, \text{attribute}), P_{\text{attribute}} \wedge P_{\text{edge}} \wedge P_{\text{join}}) \quad (4)$$

其中,式(1)表示:权限约束到 $\text{object}_{\text{G-all}}$意味着权限无条件的约束到表 Edge 和表 Attribute 中的所有字段;式(2)表示:权限约束到 $\text{object}_{\text{G-edge}}$意味着权限有条件或无条件的约束到表 Edge 的所有或部分字段;式(3)表示:权限约束到 $\text{object}_{\text{G-AV}}$意味着权限有条件或无条件的约束到表 Attribute 的所有或部分字段;式(4)表示:权限约束到 $\text{object}_{\text{G-edge\&AV}}$意味着权限有条件(条件非空)的约束到表 Attribute 和表 Edge 的部分字段.

由定义可知,$\text{object}_G \in D_G$, $\text{object}_R \in D_R$,和映射 C 相比,映射 M 增加了权限的约束条件,给出了更细粒度的映射,同时,映射 M 是映射 C 的基础上给出的,映射 M 要满足 C 的映射规则.

定义 8　如果 object_G 和 object_R 同时满足 $M(\text{object}_G) = \text{object}_R$ 和 $\text{object}_G = M^{-1}(\text{object}_R)$,则称 object_G 和 object_R 等价,记为 $\text{object}_G \equiv \text{object}_R$.

引理 1　对于 GLSDM 上任意粒度的 object_G,总存在唯一的 object_R,使得 $\text{object}_G \equiv \text{object}_R$.

证明:由 GLSDM 的结构可以看出,任意粒度的 object_G 按其权限约束涉及的对象不同,均可归结到 $\text{object}_{\text{G-all}}$、$\text{object}_{\text{G-edge}}$、$\text{object}_{\text{G-AV}}$和 $\text{object}_{\text{G-edge\&AV}}$这 4 种类型,很显然有:任意粒度的 object_G 均可以通过式(1)~(4)的组合映射到相关 object_R.又因为映射 C 和 M 均为一一映射,则得证.

引理 2　对于关系表 Edge 和 Attribute 上任意粒度的 object_R,总存在唯一的 object_G,使得 $\text{object}_G \equiv \text{object}_R$.

证明类似引理 1,此略.

定义 9　访问控制规则描述访问者是否对被访问的资源拥有某些操作权限.RDB 的访问控制规则记为 $R_R = \{\text{subject}, \text{object}_R, \text{action}, \text{sign}\}$.其中,subject 表示主体,$\text{object}_R$ 含义同定义 4,action 表示主体对资源进行的 read、update、delete 等操作,sign 是授权标识{"+","-"},"+"表示授权,"-"表示不授权.

定义 10　基于 RDB 的 GLSDM 访问控制规则记为 $R_G = \{\text{subject}, M(\text{object}_G), \text{action}, \text{sign}\}$.其中,$M$ 将 object_G 映射到 object_R,其它部分的含义同定义 9.

4.3　包含权限信息的动态查询重写

需要将 R_G 转换到 R_R 才能使用 VPD 技术,为此我们提出如下算法来解决这个问题.

算法 DQR(Dynamic Query Rewriting)

输入:用户在 GLSDM 上的查询语句 Q,该用户的访问控制规则$\{R_{G1}, \cdots, R_{Gk}\}$;

输出:包含权限信息的查询 Q'.

步骤:

1　for $i = 1$ to k do
2　　计算 $M(\text{object}_{Gi}) = \text{object}_{Ri} = (F_i, T_i, P_i)$;
3　　将 P_i 写入 VPD 策略函数(policy function)PF_i;
4　　将 R_{Gi}的其它信息写入 VPD 策略(policy):将函数 PF_i 与要保护的表、行或栏(即 F_i, T_i 信息)进行关联;如果 sign 为"-",将要保护的内容设为 null 进行屏蔽;
5　end do
6　执行 policy 将权限信息添加到 Q 中,得到 Q'.

算法 DQR 首先将 R_{Gi}的 object_{Gi}映射到 R_{Ri}的 object_{Ri},即计算 $M(\text{object}_{Gi}) = \text{object}_{Ri}$. $\forall P_i \in \text{object}_{Ri}$,构建 VPD 策略函数 PF_i, R_{Gi}的其它权限信息写入 VPD 策略,最后执行策略改写 Q.由 R_{Ri}和 R_{Gi}的定义可知,除了 object 不同,这两者其他部分的定义是一样的.因此从本质上看,算法 DQR 通过映射 M 将 R_{Gi}转换成 R_{Ri},再通过 VPD 技术用 R_{Ri}改写 Q,进而实现细粒度和动态访问控制.

4.4　安全性分析和正确性证明

定义 11　如果查询 Q 满足以下条件,则称 Q 是安全查询(Safe Query),记为 SQ:

(1) Q 访问的数据是访问控制规则允许访问的;(2) Q 没有访问被访问控制规则禁止访问的数据.

特别的,GLSDM 上的查询 Q 是安全的,只要:

$$\text{SQ}_G = Q \cap (\bigcup_{i=1}^{n} R_{Gi}^{+} - \bigcup_{j=1}^{m} R_{Gj}^{-}) \quad (5)$$

其中,R_{Gi}^{+}是 sign 为"+"的权限规则,R_{Gj}^{-}是 sign 为"-"的权限规则;这里的并、交和差等同于 SQL 的 INTERSECT、UNION 和 EXCEPT.

定义 12　由安全查询 SQ 返回的结果是安全结果(Safe Answer),记为 SA.

定理 1 算法 DQR 返回的查询 Q' 是安全查询.

证明:只需证明 $Q' = SQ_G$.算法 DQR 可形式化表示为:

$$Q' = Q \oplus \bigcup_{i=1}^{k} VPD_{Ri},$$

其中,VPD_{Ri}表示包含 R_{Ri}信息的 VPD 函数和策略,$\oplus$表示用 VPD_{Ri}改写 Q 的过程.

由定义可知,R_R 和 $object_R$ 一样可用 SQL 描述,又由 VPD 的原理[11]可知,$\oplus$过程等同于 Q 和 R_{Ri}的交的过程,则:

$$Q \oplus \bigcup_{i=1}^{k} VPD_{Ri} = Q \cap \bigcup_{i=1}^{k} R_{Ri}$$

算法 DQR 将 R_{Ri}^{-}需要保护的内容设为 null,相当于:

$$Q \cap \bigcup_{i=1}^{k} R_{Ri} = Q \cap (\bigcup_{h=1}^{n} R_{Rh}^{+} - \bigcup_{j=1}^{m} R_{Rj}^{-})$$

$$= Q \cap (\bigcup_{h=1}^{n} (subject, object_{Rh}, action, "+") - \bigcup_{j=1}^{m} (subject, object_{Rj}, action, "-"))$$

由引理 1、2 得,

$$Q \cap (\bigcup_{h=1}^{n} (subject, object_{Rh}, action, "+") - \bigcup_{j=1}^{m} (subject, object_{Rj}, action, "-"))$$

$$= Q \cap (\bigcup_{h=1}^{n} R_{Gh}^{+} - \bigcup_{j=1}^{m} R_{Gj}^{-}) = SQ_G. \quad \text{证毕}$$

查询结果是关系表形式的记为 A_R,查询结果是 GLSDM 形式的记为 A_G,查询 Q 在 D_R 上的结果记为 $Q < D_R >$,查询 Q 在 D_G 上的结果记为 $Q < D_G >$,则图 2 的访问控制框架可形式化为:

$$A_G = C^{-1}(A_R) = C^{-1}(Q' < D_R >)$$

$$= C^{-1}(Q' < C(D_G) >) \quad (6)$$

值得注意的是,算法 DQR 返回的 Q'的直接的查询结果是关系表形式的 A_R,而非 A_G.

定理 2 对于 GLSDM 上任意的权限规则 R_G 和任意的查询 Q,通过映射 C 和 M 可以找到等价的 $object_G$ 和 $object_R$,使得算法 DQR 返回的 Q'的查询结果 A_G 是安全结果.

证明:只需证明 $A_G = SA_G$.

由定理 1 和定义 12 得,算法 DQR 返回的查询 Q'是安全的,则 A_R 也是安全的:

$$A_R = Q' < C(D_G) > = Q' < D_R > = SA_R \quad (7)$$

由(6)和(7)可得:$A_G = C^{-1}(SA_R)$.

又 $SA_G \in D_G$,且 SA_G 可看作特殊的 $object_G$.由引理 1,对于任意的 $object_G$,映射 M 可找到唯一等价的 $object_R$,则由映射 M 和 C 的关系可知:$SA_G = C^{-1}(SA_R)$,即:$A_G = C^{-1}(SA_R) = SA_G$. 证毕

定理 2 说明了本文的框架是正确的:映射 C、M 和算法 DQR 能够保证对 GLSDM 的访问控制通过等价的 RDB 的访问控制来实现.

5 实验结果及分析

实验硬件环境为 Intel(R) Atom(TM) CPU 1.66GHz,内存 1 GB,硬盘 160 GB,操作系统为 Microsoft Windows XP.本文 GLSDM 描述的学术信息主要来源于我们公开的实验平台:学者网(http://www.scholat.com/).该 GLSDM 共有 857 个节点,4216 对节点的 attribute-value 和 340 条边.将 GLSDM 转换到 Oracle 11g 数据库中,得到 1 个 Edge 表和 23 个 Attribute 表.定义如下用户及其对 GLSDM 的访问规则,分别涉及关系表的 table,column,row 和 cell 安全级别:

(*a*)User1 可访问整个 GLSDM;

(*b*)User2 禁止访问节点的 phone 属性;

(*c*)User3 可访问名为 Authoredby 的边;

(*d*)User4 可访问节点的属性 email,要求这些节点是边 Authoredby 的引出节点且节点的属性 workunit 的值和 User4 相同.

用于测试的查询语句如下:

*Q*1:查询在武汉召开的会议信息;

*Q*2:查询引出节点 ID 为 77 的边的信息;

*Q*3:查询由边 Authoredby 连接的节点的信息;

*Q*4:查询节点的属性为 englishname,email 和 workunit 的信息.

结果:

	TITLE	LABLE	ENGLISH_NAME	PHONE_NUMBER
236	A Strategy for Selecting...	Authoredby	Yuxia sun	(null)
237	A Scheme for Dynamic Det...	Authoredby	Yuxia sun	(null)
238	描述逻辑 μALCIO 的语义及...	Authoredby	Yong Tang	(null)
239	SA: 一种有利于多属性范围...	Authoredby	Yong Tang	(null)
240	Mapping Bitemporal XML D...	Authoredby	Yong Tang	(null)
241	时态XML索引技术	Authoredby	Xiaoping Ye	(null)
242	Study and application of...	Authoredby	Xiaoping Ye	(null)
243	Error Bounds For Glimm D...	Authoredby	Xiaoping Ye	(null)
244	Realization of Program B...	Authoredby	Dai qingyun	(null)
245	装备制造业MES系统的设计...	Authoredby	Dai qingyun	(null)

(*a*) User2执行*Q*3

结果:

	ENGLISH_NAME	EMAIL	WORK_UNIT
63	Qi Deyu	(null)	华南理工大学
64	Jerry	(null)	东北大学
65	guoliangChen	(null)	深圳大学
66	jizhen	(null)	深圳大学
67	mingzhong	(null)	深圳大学
68	Fangwenchong	f01020126.com	中山大学
69	Liangchao YAO	yaosuperman@126.com	中山大学
70	Jun Zhu	(null)	东莞理工学院
71	anson	ansonjoe@126.com	中山大学
72	Lingkun Wu	wulingkun@qq.com	中山大学
73	Rocky	shipan@mail2.sysu.edu.cn	中山大学

(*b*) User4执行*Q*4

图4 User2执行查询*Q*3和User4执行查询*Q*4的结果

User2 执行查询 $Q3$、User4 执行查询 $Q4$ 的结果分别如图 4(a)和 4(b)所示:User2 不能查看 phone 属性,该属性的值被屏蔽,实现了 column 级安全访问;User4 只能查看和自己工作单位(此处为"中山大学")相同的学者的 email,其他单位学者的 email 被屏蔽,实现了 cell 级安全访问.

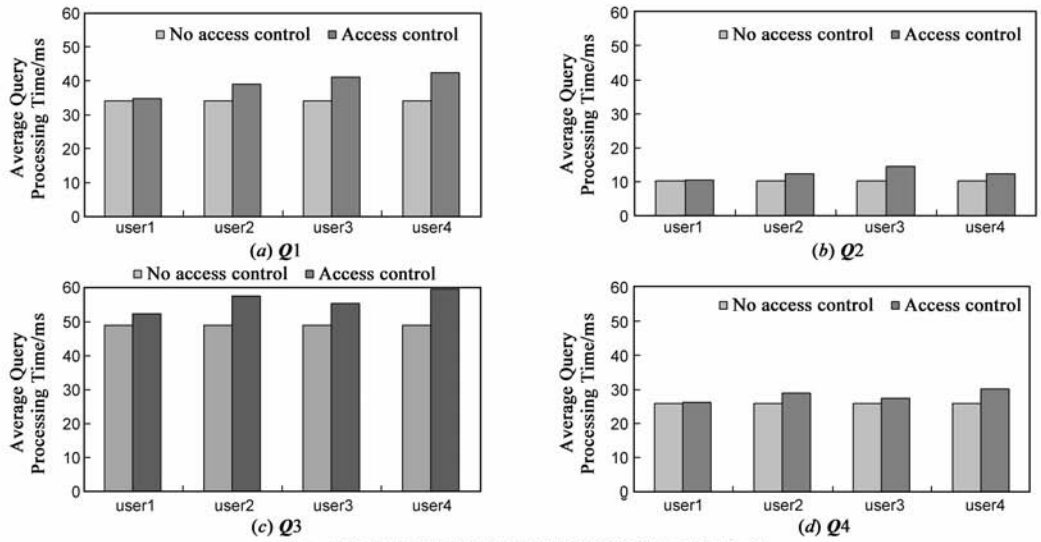

图5　无权限控制和有权限控制时查询处理时间比较

我们比较无权限控制时和使用本文的权限控制方法时查询处理时间的不同,实验结果如图 5 所示,查询处理时间取 10 次平均值.有/无权限控制时 $Q1-Q4$ 的查询处理时间相差不大,在可接受的范围.有/无权限控制时查询处理时间均较大的是 User1-User4 执行查询 $Q3$ 时(见图 5(c)),这是因为 $Q3$ 涉及较多的表连接操作,尤其是 Attribute 表间的连接,可用聚类方法将具有多个相同属性的节点组成一个表,减少 Attribute 表的个数,进而减少 Attribute 表间的连接.从总体上看,本文的方法并没有增加过多的负担,是可行的.

6　结论

本文提出一个基于 RDB 的访问控制框架,实现了极松散结构模型的细粒度和动态的访问控制.数据空间的查询语言一般为 XML 的 XPath、XQuery 查询语言或与此类似(如 iDM 的查询语言 iQL 可直接写成 XPath 和 XQuery 的形式[4]).此外,GLSDM 比 XML 更接近于关系模型的结构[15],而且本文定义的基于 RDB 的 GLSDM 访问控制规则依赖于数据模型,因此,GLSDM 访问规则到 RDB 访问规则的转换,会比 XML 访问规则到 RDB 访问规则的转换更简单.由此可知,本文方法涉及的转换代价(即数据空间查询语言、GLSDM 及其访问规则转换为 SQL、关系模型和 RDB 的访问规则所付出的代价)和基于 RDB 的 XML 文档的访问控制的转换代价[6]一样,可以控制在一个合理的范围内.我们下一步工作将研究数据空间交叉重叠时访问控制的实现问题,在这种情况下实现访问控制需要考虑更多的因素.

参考文献

[1] Franklin M, Halevy A, Maier D. From databases to dataspaces: a new abstraction for information management[J]. ACM SIGMOD Record, 2005, 34(4): 27 – 33.

[2] Halevy A, Franklin M, Maier D. Principles of dataspace systems [A]. Proceedings of 25th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems [C]. Chicago, IL, United States: Association for Computing Machinery, 2006. 1 – 9.

[3] Jin L, Zhang Y, Ye X. An extensible data model with security support for dataspace management [A]. Proceedings of 10th IEEE International Conference on High Performance Computing and Communications [C]. Dalian, China: IEEE, 2008. 556 – 563.

[4] Dittrich J P, Salles M A. iDM a unified and versatile data model for personal dataspace management [A]. Proceedings of the 32nd International Conference on Very Large Data Bases [C]. Seoul, Korea: VLDB Endowment, 2006. 367 – 378.

[5] Lee D, Lee W C, Liu P. Supporting XML security models using relational databases: a vision[A]. Proceedings of Xsym(XML Database Symposium) [C]. Berlin, Germany: Springer, 2003. 267 – 281.

[6] Luo B, Lee D, Liu P. Pragmatic XML access control using off-the-shelf RDBMS[A]. Proceedings of ESORICS (European Symposium On Research In Computer Security) [C]. Dresden, Germany: Springer, 2007. 55 – 71.

[7] Koromilas L, Chinis G, Fundulaki I, et al. Controlling access to XML documents over XML native and relational databases

[A]. Proceedings of Secure Data Management [C]. Lyon, France: Springer, 2009. 122 – 141.

[8] Lefevre K, Agrawal R, Ercegovac V, et al. Limiting disclosure in hippocratic databases [A]. Proceedings of the Thirtieth International Conference on Very Large Data Bases [C]. Toronto, Canada: VLDB Endowment, 2004. 108 – 119.

[9] Chaudhuri S, Dutta T, Sudarshan S. Fine grained authorization through predicated grants [A]. Proceedings of IEEE 23rd International Conference on Data Engineering [C]. Istanbul, Turkey: IEEE, 2007. 1174 – 1183.

[10] Barker S. Dynamic meta-level access control in SQL [A]. Proceedings of 22nd Annual IFIP WG 11.3 Working Conference on Data and Applications Security [C]. London, UK: Springer, 2008. 1 – 16.

[11] Oracle Corporation. Oracle® Database Security Guide 11g Release 2 (11.2) E10574-04 [EB/OL]. http://www.oracle.com/. 2009-11-21.

[12] 蔡伟鸿,韦岗,肖水.基于映射机制的细粒度 RBAC 委托授权模型[J].电子学报,2010,38(8):1753 – 1758.
CAI Wei-hong, WEI Gang, XIAO Shui. Fine-grained role delegation model based on mapping mechanism [J]. Acta Electronica Sinica, 2010, 38(8): 1753 – 1758. (in Chinese)

[13] 沈晴霓,卿斯汉,贺也平,等.一种支持动态调节的最小特权安全策略架构[J].电子学报,2006,34(10):1803 – 1808.
SHEN Qing-ni, QING Si-han, HE Ye-ping, et al. A framework for implementing dynamically modified least privilege security policy [J]. Acta Electronica Sinica, 2006, 34(10): 1803 – 1808. (in Chinese)

[14] Dong X, Halevy A. Indexing dataspaces [A]. Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data [C]. Beijing, China: ACM, 2007. 43 – 54.

[15] Fletcher G H, Van B J, Van G D, et al. Towards a theory of search queries [A]. Proceedings of the 12th International Conference on Database Theory [C]. Saint Petersburg, Russia: ACM, 2009. 201 – 211.

[16] Salles M, Dittrich J, Blunschi L. Intensional associations in dataspaces [A]. Proceedings of the 26th International Conference on Data Engineering [C]. Long Beach, USA: IEEE, 2010. 984 – 987.

作者简介

潘 颖 1972 年 9 月生,中山大学计算机应用技术专业博士研究生,广西师范学院计算机与信息工程学院高级实验师,主要研究方向为数据库和 Web 服务.
E-mail: panying6@sysu.edu.cn

汤 庸 1964 年生,博士,华南师范大学计算机学院教授,博士生导师,中国计算机学会高级会员,主要研究方向为数据库和知识工程.
E-mail: ytang@scnu.edu.cn

刘 海 男,1974 年 7 月出生于湖南省张家界市.博士.主要研究方向为数据库与协同软件.
E-mail: namelh@gamil.com