基于 RDF 的 XML 安全推理控制

李 专 王元珍

(华中科技大学计算机科学与技术学院 武汉 430074)

摘 要 提出了一种新的基于 RDF 的 XML 安全推理控制方法,将文档节点封装为 XML 对象,通过 XML 对象和类型刻画节点之间的语义关系,极大地拓展了推理控制范围。将节点的授权转换为对象/类型的授权,解决了面向节点授权模型难以处理的聚合敏感问题,同时也简化了面向节点方式下的繁杂授权过程。

关键词 XML 访问控制,推理控制,推理闭包

RDF Based Secure XML Inference Control

LI Zhuan WANG Yuan-Zhen

(College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074)

Abstract In this paper we present a RDF based inference control framework that provides high security and flexibility for XML documents. As objects are used to encapsulate XML nodes and to be the targets of authorization instead of the nodes, the problems of semantic description and aggregation sensitivity among nodes are resolved.

Keywords XML access control, Inference control, Inference closure

1 引言

作为未来关系型数据库的潜在替代者,XML 除了需要拓展其数据存储和交换的基本功能之外,更重要的是要保证其存储的数据的安全性。访问控制模型由于能够有效地描述安全需求,因而被迅速移植到 XML[6~8]。但访问控制只限制了直接的信息泄露,还存在由相互关联的数据引起的间接泄露。如图 1 所示,Alice 所患疾病是敏感的,由于病房与疾病相关联,通过同病房 Cathy 的疾病就可以推理出该敏感信息。显然 XML 访问控制模型不能限制这种推理行为。

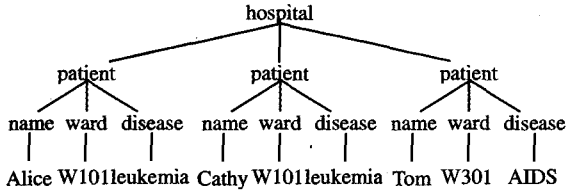


图 1 关于医院信息的 XML 文档片断

现有 XML 推理控制研究[1~3]针对的是节点之间的简单约束关系,并不能涵盖多数的语义关系。本文的推理控制技术以 RXACL[4]为基础,借助其 XML 对象的定义,通过 RDF 的强大语义描述能力来刻画对象间的关系,以如实反映 XML 文档结构无法描述的节点间的各种联系。通过 XML 类型的引入简化授权处理过程,为描述复杂的语义关系提供了基础;通过计算对象的推理闭包实现推理控制,最终将推理控制与访问控制紧密结合在一起,进一步提高 XML 文档的安全性。

2 安全访问控制

2.1 基本概念

XML 文档由嵌套的元素和值组成;元素由一对表示开始和结束的标记界定;标记可以任意。文档通常以树型结构描述,其中内节点和叶节点分别对应于文档的元素和值;边表示嵌套关系。节点可以通过路径表达式定位,如 $p:l_1/l_2/\cdots/l_n$ 确定的节点是沿着名称序列 $l_1,l_2,\cdots,l_{n-1}$ 遍历某文档树 t 得到的名为 l_n 的节点;确定的子树是以 l_n 为根的子树,用 $p(t)$ 表示。XML 文档结构只能刻画节点间的层次关系。RDF[5]最初用于描述 Web 中可以被 URI 标识的各种资源及其相互关系,借助其同样可以刻画 XML 文档节点间的语义关系。

定义 1(XML 对象) 经 RDF 陈述唯一封装、由单一路径表达式定位的 XML 文档子树称为简单对象;经 RDF 陈述唯一封装、由具有公共根节点的多个已封装对象构成的文档子树成为组合对象。简单对象和组合对象统称为 XML 对象。

例 1 以元组形式给出图 1 中的部分 XML 对象(使用这种非 RDF 规范形式是出于篇幅上的考虑。)

简单对象 *SO*01:(*SO*01, *locate*: *name/Alice*, *type*: *SimpleObject*); *SO*02:(*SO*02, *locate*: *disease/leukemia*, *type*: *SimpleObject*); *SO*03:(*SO*03, *locate*: *disease/AIDS*, *type*: *SimpleObject*)

组合对象 *AO*01:(*AO*01, *asscroot*: *hospital/patient*, *component*:{*SO*01, *SO*02}, *type*: *AsscObject*)

其中,主体 *SO*01 等表示对象唯一区分标识 OID;谓词 *type* 表示主体类型;*locate* 定位简单对象;*asscroot* 表示公共根节点;*component* 说明组合对象的构成。

节点唯一封装于一个简单对象,任意节点子集必然有唯一的简单对象集与之对应,故对象上的敏感指定可以自然转移到节点,以对象为单位的授权与以节点为单位的 XML 访问之间不存在矛盾,对象的透明性不会影响到授权的有效性。

2.2 敏感对象及其授权

实际应用中存在两种形式的敏感问题:单一敏感和聚合敏感。前者的所有组成元素的敏感性相同;对于后者,聚合体作为整体可以有与其组成元素不同的敏感性。XML 访问控制模型应该支持整体与个体的分离授权,这是面向节点模

李 专 博士研究生,主要研究方向为数据库安全。王元珍 博士生导师,教授,主要从事数据库与多媒体技术的理论及实现技术研究。

型[6~8]难以实现的。通过 XML 对象可以简单刻画上述需求，即用简单对象描述单一敏感的节点或子树；用组合对象刻画敏感聚合体。不失一般性，可以只考虑授权的主体、客体以及类型三个基本成分。

定义 2(XML 授权) 对授权的主体、类型及客体的一条 RDF 陈述。其中类型分为正授权(+)和反授权(-)，代表允许和拒绝访问。授权以 XML 对象为基本单位，对象与其组成满足约定：(1)简单对象封装的节点与对象自身的授权相同；(2)正授权组合对象的子对象同样具备正授权；(3)反授权组合对象的子对象可以有不同于整体的授权。

例 2 满足单一敏感需求“疾病 AIDS 对用户 *US*01 敏感”的 XML 授权。

XML 授权 *AU*01：(*AU*01，*user*：*US*01，*accesstype*：-，*object*：*SO*03，*type*：*Authorization*)

例 3 满足组合敏感需求“Alice 所患疾病对用户 *US*01 敏感”的 XML 授权。

XML 授权 *AU*02：(*AU*02，*user*：*US*01，*accesstype*：-，*object*：*AO*01，*type*：*Authorization*)

访问控制模型根据用户在对象上的授权检查其当前对节点的查询能否获得许可。考虑到封装的特点，我们约定：(1)简单对象中，单个节点的泄露等同于整体的泄露；(2)组合对象中，只有所有子对象的泄露才会导致整体的泄露。

2.3 授权模型的改进

相对于面向节点的访问控制模型，上述面向对象的模型可以简化敏感信息的指定。但当例 3 中的安全需求扩展到所有患者时，授权信息依然会出现大量冗余。如果从结构相同、仅叶节点不同的对象抽象出类型，则可以进一步简化授权过程，同时也可以为刻画语义关系提供基础。

定义 3(XML 类型) 简单类型的定义类似简单对象，但不包含叶节点；组合类型由简单类型和/或其它组合类型构成。类型用 CID 唯一区分。

例 4 以元组的形式给出图 1 中的部分 XML 类型。

简单类型 *SC*01：(*SC*01，*locate*：*name*，*type*：*SimpleClass*)；*SC*02：(*SC*02，*locate*：*disease*，*type*：*SimpleClass*)

组合类型 *AC*01：(*AC*01，*asscroot*：*hospital/patient*，*component*：{*SC*01，*SC*02}，*type*：*AsscClass*)

授权也可以直接在 XML 类型上进行。XML 对象是其所属类型的一个实例，具有类型的所有敏感指定；可以直接改变某对象的敏感指定，而不影响同类型的其它对象。所有授权的集合构成了 XML 文档的一个安全策略 SP。SP 中往往存在授权冲突，单一客体的授权冲突研究较多[6,7]，解决方法是在冲突发生时根据预先设定的规则自动确定优先权限，并且在规则不变条件下结果是恒定的。不同客体间的授权冲突发生在语义相关的对象之间，是引发推理泄露的根源，推理控制的实质就是要解决这种冲突。

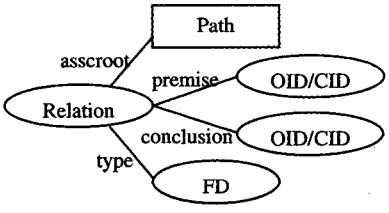


图 2 对象关系的 RDF 表示

节点对象化、对象类型化之后，通过 RDF 可以方便地描述上述语义关系。图 2 给出的是函数依赖型关系的 RDF 图示，谓词 asscroot 用于确定关系双方的公共根节点；type 对应的客体 FD 作为资源可以进一步在系统实现时使用 RDF 来刻画其在推理控制过程中的处理策略。

例 5 以元组的形式给出图 1 中的一个 XML 函数依赖关系。

简单类型 *SC*04：(*SC*04，*locate*：*ward*，*type*：*SimpleClass*)

XML 关系 *R*01：(*R*01，*asscroot*：*hospital/patient*，*premise*：*SC*04，*conclusion*：*SC*02，*type*：*FD*)。

3 安全推理控制

3.1 推理控制假定

用户行为的任意性使得引起推理泄露的途径各式各样，但并不都需要控制。参照关系数据库安全研究中的封闭世界假设(CWA)[9]，可以给出类似的假定，以限定 XML 推理控制问题的解决范围。

假定 1 推理依赖于 XML 文档中的节点及相互关系。与 XML 文档无任何联系的“凭空”推理是难以控制的。

假定 2 推理结果应存在于 XML 文档中。只有造成文档中敏感数据泄露的推理才需要控制。

在封闭式访问控制策略下，某用户在给定文档上的所有授权构成了对该文档的一个分划：允许集、显式拒绝集和隐式拒绝集。为平衡数据安全性与其可用性之间的矛盾，我们约定：访问控制保证所有可访问的对象都经过了许可；推理控制确保显式拒绝集中的对象不被推理，对于隐式拒绝访问的对象，即使可被推理也不需要控制。

3.2 推理扩展

推理是一个反复过程，新加入的对象可能激活其它对象关系用于进一步的推理。若所有对象关系作用在某对象集上都不再推理出新对象，则称此时的对象集是推理完备的。

定义 4(推理闭包) 将某 XML 文档树 t 的所有对象关系作用在其对象集 O 上得到的一个推理完备集，用 O^+ 表示。由于 t 中推理关系是有限的，根据前述假定，将其作用于 O 上得到的推理结果显然也是有限的，因此 O^+ 是确定存在的。

推理闭包用于判定某允许集是否存在推理泄露，即给定用户 u 在文档 t 中的可访问对象集 O，若推理闭包 O^+ 中不存在 u 的显式反授权对象，则 t 对 u 是推理安全的。

算法：推理闭包的计算

输入：文档 t 的目标对象集 O，所有组合对象集 O_c，对象关系集 R

输出：推理闭包 O^+

方法：

1. 令 O' 表示当前处理结果集，$O'=O$

2. 若 $R=\varnothing$，则返回 O'，否则从 O 中任取一对象 o，

(a)取 R 中 o 对应的关系子集 R'，获取相应推理结果集 O''，若 O'' 中存在组合对象，则将其所有子对象亦并入 O'' 中

(b)$R=R-R'$，$O=O\cup(O''-O')$，$O=O-\{o\}$，$O'=O'\cup O''\cup\{o\}$

(c)若 $O\neq\varnothing$，则重复执行 2

3. $O_c=O_c-O'$，对 $\forall o'\in O_c$，若其所有子对象均属于 O'，则 $O=O\cup\{o'\}$，重复执行 2，否则返回 O'

推理通常并不是通过一次查询完成，保留用户历史查询

(下转第 139 页)

考察以上算法在实际使用中的效率性，以及探讨超协调本体推理在非单调推理的推广。

参 考 文 献

1 Schlobach S, Cornet R. Non-standard reasoning services for the debugging of description logic terminologies. In: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'03), Acapulco, Mexico, August 2003. 355～362

2 董明楷，蒋运承，史忠植．一种带缺省推理的描述逻辑．计算机学报，2003，26(6)：729～736

3 Huang Z, Harmelen F, Teije A. Reasoning with Inconsistent Ontologies. In: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI'05), Edinburgh, Scotland, August 2005. 454～459

4 Benferhat S, Garcia L. Handling locally stratified inconsistent knowledge bases. Studia Logica, 2002. 77～104

5 Lang J, Marquis P. Removing inconsistencies in assumption-based theories through knowledge-gathering actions. Studia Logica, 2001. 179～214

6 Marquis P, Porquet N. Resource-bounded paraconsistent inference. Annals of Mathematics and Artificial Intelligence, 2003. 349～384

(上接第 105 页)

结果也是推理控制的关键。将节点封装为对象后，只需在历史记录中记载已访问对象的标识，这样可以减少存储负担。

3.3 推理控制

推理控制可以在设计和查询两个阶段实施。设计阶段可以根据敏感对象的反向推理链，确定引发推理泄露的对象，进而通过授权调整来消除泄露。但是，授权调整是以牺牲数据可用性为代价的，对于聚合敏感问题，基于信息最大可用的原因，授权调整并不能有效解决，只能在查询阶段实施控制，这样才能在保证文档安全同时，提高数据可用性。

过程：查询阶段的推理控制

输入：用户 u 的历史记录 O，显式拒绝集 O'，当前查询结果 O'，对象关系集 R

输出：允许或拒绝查询，新的历史记录

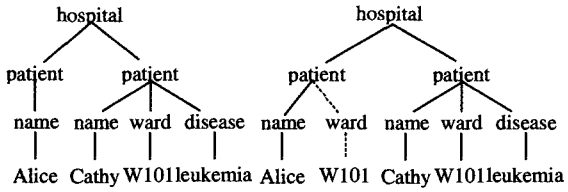
方法：

1. $O_1=O\cup O'$，根据算法 1，根据 R 计算 O_1^+

2. 检查推理泄露：对 $\forall o\in O'$，若 $o\in O_1^+$，则拒绝查询并返回 O

3. 允许查询并返回 O_1

例 6 仅在授权 $AU02$ 下针对用户 $US01$ 的安全推理控制。



(a) 已访问的文档内容　　(b) 因推理泄露而拒绝访问的文档内容

图 3　XML 文档安全推理控制示例

图 3(a)是 $US01$ 当前符合授权策略的已访问内容，对象 $SO01$ 包含其中；当其试图访问(b)中虚线指示的节点时(即对象 $SO04$：($SO04$, *locate*: *ward*/$W101$, *type*: *SimpleObject*))，使用上述过程根据关系 $R01$，可以推理出与 $SO01$ 具有公共根节点的对象 $SO02$，从而得到敏感的组合对象 $AO01$。故控制系统将拒绝 $US01$ 对 *Alice* 所属病房信息的查询，保证敏感信息不被推理泄露。

4 系统框架及实现

基于 RDF 的 XML 文档推理控制的体系结构如图 4(1～9 为运行阶段控制流程；I～II 为安全设计)所示，主要由三大模块组成，分别为查询引擎、访问控制和推理控制。

我们在达梦 SDM4 数据库管理系统的 XML 数据库上实现上述系统。该数据库采用面向节点授权的访问控制机制，支持满足 XPath 规范的路径查询。推理控制系统在此基础上实施。在设计阶段，完成节点封装、对象关系指定和授权，最终形成给定 XML 文档的安全策略。在查询阶段，用户提交的查询经由查询引擎得到查询结果，先由访问控制模块根据安全策略检查授权合法性，若违反则拒绝响应，否则将结果提交给推理控制模块。后者先将结果与历史记录进行临时归并，然后利用对象关系对归并结果进行推理扩展，检查是否存在推理泄露以决定对用户的响应。如果不存在推理泄露，在响应用户之前还需更新原有历史记录。

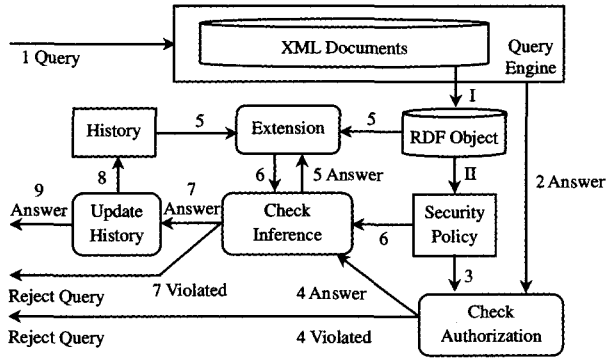


图 4　XML 推理控制系统体系结构

总结　本文在 RXACL 的基础上进一步通过 RDF 刻画 XML 文档中节点间的语义关系，提出了基于 RDF 的 XML 安全推理控制方法。通过 XML 类型的引入扩展了 RXACL 模型，简化了复杂的授权过程；以对象标识形式存储的历史记录可以降低存储负担，同时也避免了复杂的文档归并过程。后续的研究将致力于解决文档间的关联推理问题。

参 考 文 献

1 Yang XiaoChun, Li Chen. Secure XML Publishing without Information Leakage in the Presence of Data Inference. In: Proc. of the 30th VLDB Conference, 2004. 96～107

2 Tan Z J, Pang Y M, Shi B L. Reasoning about functional dependency for XML. Journal of Software, 2003, 14(9): 1564～1570

3 Farkas C, Stoica A. Correlated data inference in ontology guided XML security engine. In: Proc. of the 17th WG 11. 3 working conference on Data and Application Security, 2003

4 Gowadia V, Farkas C. RDF metadata for XML access control. In: Proc. of the 2003 ACM workshop on XML security, 2003. 39～48

5 Consortium W W W. Primer RD F. W3C Recommendation. Available at: http://www. w3. org/TR/2004/REC-rdf-primer-20040210, 2004

6 Damiani E, Vimercati S D C, Paraboschi S, et al. A fine-grained access control system for XML documents. ACM TISSEC, 2002, 5(2): 169～202

7 Bouganim L, Ngoc F D, Pucheral P. Client-Based Access Control Management for XML documents. In: Proc. of the 30th VLDB Conference, 2004. 84～95

8 Fan Wenfei, Chan Chee-Yong, Garofalakis M. Secure XML Querying with Security Views. In: Proc. of the 2004 ACM SIGMOD international conference on Management of data, 2004. 587～598

9 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, 1986