一种基于 XML Schema 的安全访问控制策略

王战敏,崔杜武
WANG Zhan-min,CUI Du-wu

西安理工大学 计算机科学与工程学院,西安 710048
School of Computer Science and Engineering,Xi'an University of Technology,Xi'an 710048,China

WANG Zhan-min,CUI Du-wu.Access control strategy based on RBAC for XML security.Computer Engineering and Applications,2007,43(17):157-159.

Abstract: In this paper,a Role-Based Access Control(RBAC) model based on XML schema is proposed for solving security problems of accessing XML documents.XML schema supports complex constraints for XML components and provides a mechanism to build rich reuse relationships between schemas and elements.Based on these features our model extends the RBAC96, corresponding algorithm-TDACA(Target Document Access Control Algorithm) is also presented,which provides a fine-grained access control and also supports instances-level authorization methods.

Key words: XML Schema;SRBAC;security;access control strategy

摘 要:为提高 XML 文档资源访问控制的安全性,依据 RBAC 对 XML 组件具有复杂约束以及模式与元素之间关系的良好重用机制的特性,构建了一种新的访问控制模型—SRBAC。在此基础上,提出了相应目标访问控制算法 TDACA,创建了相应策略,验证了 SRBAC 的有效性,从而实现了有效的访问控制和实例级的认证,保证了 XML 文件的安全使用。

关键词:XML Schema;SRBAC;安全;访问控制策略
文章编号:1002-8331(2007)17-0157-03 **文献标识码**:A **中图分类号**:TP311

1 引言

作为一种数据描述语言,XML 自身具备的可扩展特性,使得用户可以根据文档的内容定义自己的 XML 标记语言。XML Schema 是一种常用的描述信息结构的模式,用来定义 XML 文档的文本结构、数据类型等规则,为一类文档建立一个模式,规范文档中的标签(tag)和文本(text)可能的组合形式。规范的模式,为 XML 文档提供了统一的结构。

基于角色的访问控制方法(RBAC),是传统的自主访问控制(DAC)和强制型访问控制(MAC)策略之后一种新的访问控制技术,是解决大型企业的统一资源访问控制的有效方法[1]。由 Sandu 提出的 RBAC96 策略涉及用户(Users),角色(Roles),权限(Permissions)3 个部分。角色和权限之间存在的权限分配,满足多对多的映射关系,但此类映射用于控制访问 XML 文档时存在着遗失部分不符合既定模式的资源文件问题。而现今 XML 以其规范的结构广泛地应用于企业内部组织和企业间的电子商务,对数据资源的安全使用和共享提出了更高的要求。在此基础上,本文从角色访问控制角度提出了一种新的访问控制策略 SRBAC,利用模式本身对 XML 组件具有复杂约束,和对模式与元素之间关系的良好重用机制,实现有效的访问控制和实例级的认证,提供了保证 XML 文件使用安全的一种新方式,并结合实际的企业 XML 应用对策略做了进一步的说明。

2 SRBAC 模型

目前存在着多种访问控制技术,如访问控制列表(ACL)、基于角色的访问控制(RBAC)等。RBAC96[2]模型由 Barka 和 Sandu 在 1996 年提出,引入角色中介,实现了权限与角色、角色与用户相关联,从而达到了用户与访问权限逻辑分离的目的。

上述策略应用于规范的 XML 文档中,具有良好的安全访问控制能力,但典型的信息资源,总包含大量的模式文档和不受模式束缚的自由文档,对此类资源 RBAC96 访问过程复杂,且自由文档安全访问不能保证。针对此类情况,提出了一种新的 RBAC 访问控制模型—SRBAC,模式架构如图 1 所示。

该模型通过用模式权限(SP)及角色-模式权限的委派(EPA)取代角色到权限的直接委派,达到了权限划分细化,保证自由文档安全性的目的。

2.1 基本概念

定义 1[2] 用户指与计算机系统直接进行交互的任何人;角色用于组织大量用户的权限规范;用户分配即用户与角色之间多对多的关系。

定义 2 模式对象 SO(Schema Objects):一个 XML 模式或模式中的元素,也可以是 Xpath 或 Xquery 表达式。

定义 3 实例对象 IO(Instance Objects):一个 XML 实例或

基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No.40537031,No.40375010,No.60278019)。
作者简介:王战敏(1969-),女,博士生,讲师,主要研究方向:Internet 应用研究,多媒体技术与应用;崔杜武,博士生导师,教授,主要研究领域:人工智能、多媒体技术、Internet 应用研究等。

实例中的元素。

定义 4 实例映射 IM(Instance Mapping):SO 到 IO 的一对多映射。

IM:$SO\rightarrow 2^{IO}$, 即 $\exists so_1,so_2\in SO,io\in IO,(so_1\neq so_2)\wedge(io\in IM(so_1)\wedge io\in IM(so_2))$

在 SRBAC 中, 为模式对象定义的权限可以由 IM 转换成实例对象的,所以 IM 从实例层次暗示了认证的规则。

定义 5 模式对象层次关系 SSOH (Secure Schema Object Hierarchy):$SSOH\subseteq SO\times SO$,模式对象的偏序集。

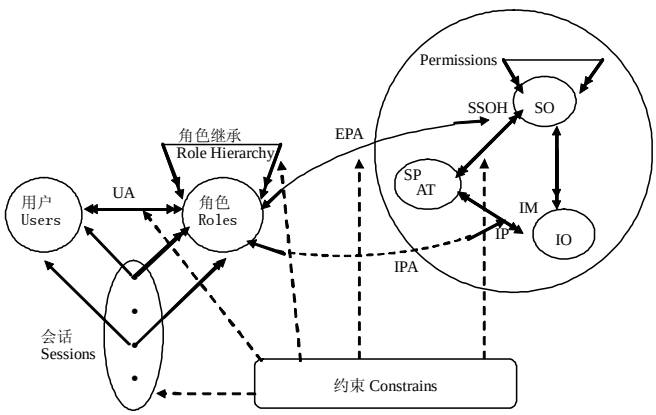

图 1 SRBAC

2.2 访问类型[3]

对 XML Object 的操作主要有四个基本类型。用户激活角色读取 XML 实例的元素、内容、属性等,为读(Read)的过程;对 XML Object 的内容、属性值等进行修改,为更新(Update),修改过的实例仍然要符合已有模式;文档编辑者在目标文档的根元素下,创建新的树结构即为创建(Create);移除 XML 实例或者实例中的元素、属性、内容等操作,称为删除(Delete)。

以下通过一个应用实例分步介绍其实现过程。某企业具有人事、财务、生产等部门,每个部门有各自不同的 XML 应用系统,在企业应用集成的过程中,要实现部门与部门之间部分资料的互相查看,部门内部文档的操作控制。

在上述实例中, 用户指代企业中的员工 id 或者某客户端的 IP 地址,角色分为全局角色和局部角色两类,将各部门分设为全局角色(如人事部、财务部等),只拥有读权限;局部角色为每个部门内部划分的角色(如组长、普通员工等),继承于全局角色,同时也自动继承了它的权限,并根据实际情况对权限进行扩展或阻塞。用户只能委派给局部角色。

SSOH 多数是以模式组件的重用为基础, 在 XML 模式中也定义了一些不同模式间数据类型、元素等的重用标准。设已知客户信息在每个目标文档中都将是相同的, 则创建一个 Customer 对象,每当需要某个客户出现在 XML 文档中时(如作为发票或客户列表的一部分), 就可以使用相同的代码来构建必要的元素。该方法降低了代码冗余,并且极大地简化了故障诊断和升级工作。

2.3 权限

权限分为模式权限(Schema-based Permissions)和实例权限(Instance-based Permissions)两种,即 $P=SP\cup IP$。

定义 6 模式权限 SP:$SP\subseteq AT\times SO$,是模式对象与其允许的访问类型的关系集;

定义 7 实例权限 IP:$IP\subseteq AT\times IO$, 是实例对象与其允许的访问类型的关系集;

定义 8 权限重用 在 SSOH 中,$\forall so_1,so_2\in SO,at\in AT$,$(so_1\leqslant so_2)\wedge((at,so_1)\in SP)\Rightarrow(at,so_2)\in SP$

权限重用主要是权限的继承, 若模式对象 so_2 继承 so_1,则 so_1 定义的权限也传递给了 so_2。此方式减少了需要定义权限的数量和复杂性。

2.4 权限委派

将原来的权限到角色的直接委派分为 EPA 和 IPA,即 $PA=EPA\cup IPA$。

定义 9 模式权限委派 EPA (Explicit Role-Permission Assignment):$EPA\subseteq R\times SP$,模式权限到用户的委派关系。

定义 10 实例权限委派 IPA(Implicit Role-Permission Assignment):$IPA\subseteq R\times IP$,实例权限到用户的委派关系。

$IPA=\{(r:R,ip:IP)|\exists(at:AT,so:SO,io:IO)\cdot[(r,(at,so))\in EPA]\wedge[ip=(r,(at,io))]\wedge[io\in IM(so)]\}$

在前例中,为每个 SO 的模式组件均加<permission>标签来标示权限,设其有两个属性:<role>,<access>,它们的值分别是角色名和已定义好的访问类型,依定义可知,任何没有指定的角色则对此对象没有访问的权限。

2.5 约束

权限的传递,是指模式元素的权限自动传递给它的所有子元素。例如,一个元素有“读”权限,则它的所有子元素和属性都具有此权限。虽然模式对象的数据是相互嵌套的,但是在一些层次上还是可以实现的。

传递的方向,在特殊的情况下,元素的权限可以传递给它的父元素。例如角色 CSR(Customer Service Representative)对元素 Customer 有 Read 的权限,而 Customer 的内容相比它的父元素更加需要得到关注, 所以 CSR 对每个包含此元素的模式或其它元素都有 Read 的权限。

系统设计者和安全管理员根据实际服务和业务逻辑来定义不同的约束。如在财务部门,单个用户不能同时拥有两个冲突的角色。

3 访问控制策略

模型建立的目的是实现安全访问控制, 即实现用户对 XML 对象执行操作的控制,流程如图 2 所示。

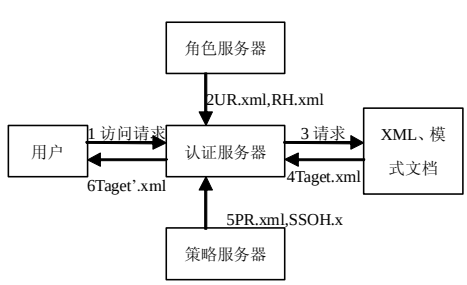

图 2 访问控制过程

用户向验证服务器发出访问目标文档请求,包括用户信息(u)、角色信息(r)、访问类型(at)、访问的目标 XML 文档(target.xml),要求目标文档满足某种模式,则应输入期望的目标模式(target'.xsd);认证服务器验证其为合法用户后,从角色服务器取得相关的角色信息并激活角色,对目标文档进行操作;策略服务器提供相关安全信息:角色-权限的委派(PR.xml)、模式对象层次关系 (SSOH.xml), 对目标文档的各组件进行提取和限制,结束返回给用户理想的目标文档。

上述控制过程中,策略服务器对目标文档的处理过程如算法 TDACA(Target Document Access Control Algorithm)所描述:

```
Input:Access request:(u,r,at,target.xml)
      Schema of target:target.xsd
      Schema of expected output:target'.xsd
      Security information:PR.xml,UR.xml,
                           RH.xml,SSOH.xml
Output:target'.xml
Method:
//验证用户-角色委派:
(1)while r∉roles(u,UR.xml,RH.xml)
(2)  ACCESS denied;
//解释生成 target.xml 实例树 t
(3)t=parse(target.xml)
(4)recur_ac(t,rootnode)
(5)so=sm(rootnode)
(6)if so∈P(at,r,PR.xml,RH.xml,SSOH.xml)
(7)  ACCESS rootnode is permited;
(8)  add(target'.xml,rootnode);
(9)  while(rootnode is not leaf)
(10)    foreach subtree st∈t rooted in subnode
(11)    Recur_ac(st,subnode);
(12)else
(13)    ACCESS t is denied;
// 验证 target'.xml 是否符合 target'.xsd ;
(14)  if target'.xml∈im(target'.xsd)
(15)   Output target'.xml;
(16)  else Access target.xml denied;
```

TDACA 以深度优先的递归算法 recur_ac(*t*,rootnode)为基础,根据目标文档中的授权信息,对文档的 DOM 树进行删减,生成符合用户身份的 XML 文档。如某用户对 XML 文档请求读操作,用户通过认证服务器的身份验证后,首先由函数 *parse*() 得到目标文档 target.xml 树结构的数据集, 解释生成实例树 *t*; *sm*(rootnode)返回实例根节点的模式对象,检查此根结点的权限,如果访问被允许,则添加至 taget'.xml 中,同时所有的子树均采用相同的检测机制,否则,访问被否决。最后返回给用户满意的 target'.xml 文档。

上文主要是针对遵循某一模式的 XML 实例进行访问控制, 直接委派给模式对象的权限间接地为实例对象委派了权限。不失一般性,特殊文档不存在固定模式,同一个模式的两个实例验证方式也存在不同, 这些均需要定义实例级的权限。XML 的良好结构, 满足权限标签不作任何修改就可直接指定给实例对象,从而实现对自由文档的访问控制。

4 结论

本文根据 XML 资源的广泛应用以及存在的访问控制问题,提出了一种适用于 XML 文档的访问控制策略。与现有模型相比,SRBAC 侧重于用标签来定义角色和权限,请求的提交和响应均采用 XML 格式。通过层次性地表示数据间关系,提高了应用的可读性和可扩展性; 权限的实例化使得对 XML 资源信息的安全检测更加周密;RBAC 自身的策略无关性和实用性加强了 SRBAC 的可操作性和灵活性。

实践过程中, 需要用户自己定义适合本领域的安全策略,定义众多的角色和访问权限及它们之间的关系,此操作复杂性较高, 同时应用中发现没有指定权限的文档无法被系统检测到, 所以提供简单实用的 XML 文档处理工具也是必不可少的工作,对此,我们将在以后的研究中逐步完善。

(收稿日期:2007 年 3 月)

参考文献:

[1] Sandhu R,Coyne E J,Feinstein H L.Role based access control models[J].IEEE Computer,1996,29(2):38-47.

[2] Osborn S,Sandhu R,Munawer Q.Configuring role-based access control to enforce mandatory and discretionary access control policies[J].ACM TISSEC,2000,3(2):85-106.

[3] Chandramouli R.Specification and validation of enterprise access control data for conformance to model and policy constraints[C]// 7th World Multi-conference on Systemics,Cybernetics and Informatics,2003.

[4] He Qing-feng.Privacy enforcement with an extended role-based access control model[R].NCSU Computer Science Technical Report,2003.

(上接 156 页)

IEEE International Conference on Information Reuse and Integration.Nevada,USA,2003:273-277.

[5] Lu Jian-jiang,Xu Bao-wen,Jiang Ji-xiang,et al.Non-negative matrix factorization for filtering Chinese document [J].Lecture Notes in Computer Science,2004,3037:113-120.

[6] Lu Jian-jiang,Xu Bao-wen,Jiang Ji-xiang.Generating different semantic spaces for document classification[J].Lecture Notes in Computer Science,2004,3309:430-436.

[7] Lu Jian-jiang,Xu Bao-wen,Huang Gang-shi,et al.Matrix dimensionality reduction for mining typical user profiles[J].Journal of Southeast University,2003,19(3):231-235.

[8] Lu Jian-jiang,Xu Bao-wen,Yang Hong-ji.Matrix dimensionality reduction for mining Web access logs[C]//The 2003 IEEE/WIC International Conference on Web Intelligence,Halifax,Canada,2003:405-408.

[9] Jiang Ji-xiang,Xu Bao-wen,Lu Jian-jiang,et al.Local nonnegative matrix factorization for mining typical user session profile[J].Lecture Notes in Computer Science,2004,3140:558-562.

[10] Berry M,Dumais T,O'Brien G.Using linear algebra for intelligent information retrieval[J].SIAM Review,1995,37(4):573-595.

[11] Inderjit S,Dharmendra S.Concept decompositions for large sparse text using clustering[J].Machine Learning,2001,42(1):143-175.

[12] Brants T,Chen F,Tsochantaridis I.Topic-based document segmentation with probabilistic latent semantic analysis[C]//The Eleventh International Conference on Information and Knowledge Management.McLean,Virginia,2002-09:211-218.

[13] Golub G,Reinsch C.Handbook for matrix computation II[M].Linear Algebra,New york:Springer Verlag,1971.

[14] Hanm J,Kamber M.Data mining:concepts and techniques[M].[S.l.]:Morgan Kaufmann Publishers,2000.