一种基于本体相似度的多域策略集成方法

金 莉 卢正鼎
(华中科技大学计算机科学与技术学院 武汉 430074)

摘 要 对各成员域的访问控制策略进行语义提取,是避免多域策略集成时可能产生概念和逻辑关系冲突的有效途径。提出一种基于本体相似度的多域互操作策略集成方法SPIOS,通过对成员域访问控制策略本体进行语义级映射,融入基于Bayesian概率的机器学习机制,自适应地归纳出能较好满足各成员域自治性和协同性的多域安全互操作策略模型。

关键词 多域,本体相似度,策略集成

Secure Multi-domain Policy Integration Method Based on Ontology Similarity

JIN Li LU Zheng-ding
(College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract Extracting semantic meaning of local access control policies is an effective method to avoid conceptual and logical conflicts in multi-domain policies integration. In view of domain ontology, a secure policy integration method based on ontology similarity was proposed. Using a machine learning algorithm of Bayesian, it can self-adaptively construct a secure multi-domain interoperation model to satisfy the autonomy and cooperation of all domains.

Keywords Multi-domain, Ontology similarity, Policy integration

多域环境下,各成员域为了满足各自的安全需求,往往制定不同的访问控制策略体系来维护本域的安全性和自治性[1,2]。由于各成员域使用的模型、语法、计划模式、数据标记模式和约束各不相同,因此在构造一条能满足各成员域安全需求和自治需求的多域互操作策略时,必须首先将成员域的访问控制策略从本地化的表达方式和术语中提取出来,再进行客观抽象的形式化描述,从而尽量避免多域策略集成时可能产生的概念和逻辑关系的冲突。

本体是对客观存在的概念和关系的描述,作为一种能在语义层上描述概念模型的建模工具,在基于RBAC的访问控制策略集成方面具有很多优势,近年来被许多学者应用于多域安全互操作策略集成的研究。M. Dzbor等人[3]提出了一个基于本体的基础访问控制模型,描述了基于授权的访问控制策略(Authority based Access Control, ABAC)。L. Kagal等人[4]提出了一种基于本体的分布式系统Rein,通过本体来实现访问控制策略的集成和共享。与上述两者借助本体共享访问控制策略不同,W. Di等人[5]将RBAC全部用描述逻辑(Description Logic)表示出来,其访问控制模型中只包含了一个描述逻辑推理器,用来构造角色层次、静态和动态的职责分离机制以及一些集合约束等,并没有进一步用其他语言来丰富访问控制系统。

一个完备的本体由类、概念、属性、关联、实例等多个部分构成,RBAC访问控制模型可以通过本体来构造用户、角色、权限、资源等概念(Concepts),定义用户、角色、权限的属性(Properties),描述角色与权限、角色与角色、权限与资源间的关联(Relations),列举角色中读取角色、写入角色等实例(Instances)。可以说,RBAC模型中的所有组件都可以通过本体给出形式化的描述和定义。因此,在构造多域互操作策略时,可以通过对各成员域访问控制策略本体进行基于语义的映射,从而提取出具有较普遍适用性的多域互操作策略,以满足多域环境下的安全性和自治性规则。

本文提出一种基于本体相似度的多域安全互操作策略集成方法(Secure Policy Integration based on Ontology Similarity, SPIOS),通过对成员域访问控制策略本体进行语义级映射(如图1所示),融入基于Bayesian概率的机器学习机制,自适应地归纳出能满足各成员域的安全性和自治性需求的多域安全互操作策略模型。

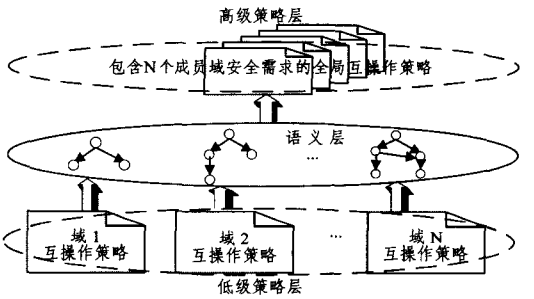

图1 基于语义的多域互操作策略集成

到稿日期:2009-04-24 返修日期:2009-07-13 本文受国家自然科学基金项目(60803114,60773191),国家高技术研究发展计划(863计划)项目(2007AA01Z403)资助。

金 莉(1978—),女,博士生,主要研究方向为信息安全、分布式计算等,E-mail:jessiewelcome@126.com;卢正鼎(1944—),男,教授,博士生导师,主要研究方向为信息安全、并行分布式处理。

1 构造 RBAC 本体

构造 RBAC 本体必须完全反映 RBAC 的原有约束和规则：用户通过获取角色来取得相应的权限，从而访问所需资源；用户在会话中激活角色，并且获得所激活角色的所有权限。因此，RBAC 本体至少包含以下 5 类概念(Concepts)。

- 用户(Users)：访问控制请求的发起者，请求对资源的访问控制权限，RBAC 模型中的主体对象；
- 角色(Roles)：RBAC 模型的核心部分，是用户与请求权限之间的主要关联，用户只能通过角色获取权限，而不能直接获取权限；
- 权限(Privilege)：表示对资源的访问控制操作；
- 资源(Resource)：RBAC 模型保护的对象客体，只有通过认证的用户才能获得对资源操作的权限；
- 会话(Session)：用户在会话中激活角色，从而获得相应的权限。

以上概念具有的属性(Properties)反映了 RBAC 中具有以下关联(Relations)。

- $hasRole \subset (User, Role)$：用户(User)获得角色(Role)；
- $activeRole \subset (User \times Session, Role)$：用户(User)在会话(Session)中激活角色(Role)；
- $hasPrivilege \subset (Role, Privilege)$：角色(Role)具有权限(Privilege)；
- $accessResource \subset (Privilege, Resource)$：权限(Privilege)可以访问资源(Resource)。

2 本体相似度映射规则

定义 1(本体[6]) 本体 O 可以表示成一个由多个属性组成的元组：

$$O_i = (C, H_C, R_C, H_R, I, R_I, A)$$

其中，C 表示概念，H_C 表示概念的层次结构，R_C 表示概念间的关联，H_R 表示关联间的层次结构，I 表示特定概念的实例，R_I 表示实例间的关联，A 表示公理。

定义 2(本体映射) 给定两个本体 A 和 B，对于本体 A 中的每一个节点(概念、关联、实例等)，都能在本体 B 中找到一个语义相同或相似的相应节点(概念、关联、实例等)，反之亦然。

从以上定义可以看出，本体间的映射可以通过逐步对本体中各节点的语义映射来完成。对本体进行语义映射，实际是对本体中概念、关联、实例等元素逐一进行相似度的比较。假设存在 O_A 和 O_B 两个本体，e_{Ai} 和 e_{Bj} 分别表示 O_A 和 O_B 中的概念、关联和实例等主要元素，那么 O_A 和 O_B 间映射的形式化描述可以用以下形式表示：

$$\begin{aligned}\mathrm{Map}(O_A \to O_B) := \{(e_{Ai}, e_{Bj}) \mid \forall e_{Ai} \in O_A, \exists e_{Bj} \in O_B,\\ \mathrm{sim}(e_{Ai}, e_{Bj}) > t(t \text{ 是临界参数})\}\end{aligned}$$

因此，本体间映射可以通过计算本体间的相似度来实现。当相似度大于某临界参数时，说明两个本体语义上是相同的，具有相应的映射关系。由于本节主要研究的是一对一的同级比较(例如概念到概念的映射)，暂不涉及跨结构的比较(例如概念到关联的映射)。

定义 3(Jaccard 相似度算法[7]) 假设 A, B 是两个进行比较的主体，那么 A 与 B 的 Jaccard 相似度可表示成如下形式：

$$\begin{aligned}\text{Jaccard-sim}(A,B) &= \frac{P(A \cap B)}{P(A \cup B)}\\ &= \frac{P(A,B)}{P(A,B)+P(A,\overline{B})+P(\overline{A},B)} \quad (1)\end{aligned}$$

由此可知，对于本体 O_A 中元素 $e_{Ai} \in \{C_A, R_A, I_A\}$，$O_B$ 中元素 $e_{Bj} \in \{C_B, R_B, I_B\}$，$e_{Ai}$ 和 e_{Bj} 间的 Jaccard 相似度可以表示成如下形式：

$$\text{Jaccard-sim}(e_{Ai}, e_{Bj}) = \frac{P(e_{Ai}, e_{Bj})}{P(e_{Ai}, e_{Bj})+P(e_{Ai}, \overline{e_{Bj}})+P(\overline{e_{Ai}}, e_{Bj})} \quad (2)$$

其中，概率 $P(e_{Ai}, e_{Bj})$，$P(e_{Ai}, \overline{e_{Bj}})$ 和 $P(\overline{e_{Ai}}, e_{Bj})$ 可以通过分别统计属于 $e_{Ai}, \overline{e_{Ai}}, e_{Bj}$ 和 $\overline{e_{Bj}}$ 的实例数与本体 O_A 和 O_B 的实例总数的比值获得。以 $P(e_{Ai}, e_{Bj})$ 为例，计算方法如下：

$$P(e_{Ai}, e_{Bj}) = \frac{O_A \text{ 中属于 } e_{Ai} \cup e_{Bj} \text{ 的实例数} + O_B \text{ 属于 } e_{Ai} \cup e_{Bj} \text{ 的实例数}}{O_A \text{ 的实例总数} + O_B \text{ 的实例总数}} \quad (3)$$

假设 t 为预先定义的临界参数，那么当 Jaccard-sim$(e_{Ai}, e_{Bj}) > t$ 时，则认为 e_{Ai} 和 e_{Bj} 语义相同，具有映射关系。

3 基于 Bayesian 学习的策略集成方法

通过 Jaccard-sim(e_{Ai}, e_{Bj}) 获得了两个本体中所有对应概念、关联、实例等两两语义映射的结果后，对于两个本体整体化的语义映射而言，各组成部分语义映射的结果最终可以影响本体语义映射的结果。采用基于 Bayesian 的机器学习算法对本体的相似度进行预测，将各元素语义映射的结果作为已知事件构成了 Bayesian 算法的训练集。

3.1 Bayesian 学习算法

Bayesian 学习算法是观察者对某一事件发生的相信程度，根据先验知识和统计数据可推测出未知事件发生的可能性，具体描述如下：

样本集合 $S=\{X_1, X_2, \cdots, X_n\}$，对于定义在样本集合 S 上的属性域 $A=\{A_1, A_2, \cdots, A_m\}$，样本 X_i 可以表示为 $X_i=\{x_1, x_2, \cdots, x_m\}$，其中 x_i 与属性域中 A_i 相对应。Y 为给定假设事件，$P(Y)$ 为其客观概率，S 为已知样本集合，$P(Y|S)$ 为后验概率。那么根据 Bayesian 理论，有：

$$P(Y|S) = \frac{P(S|Y)P(Y)}{P(S)} \quad (4)$$

假设 Y 可以表示成 $Y=\{y_1, y_2, \cdots, y_m\}$，且 m 个属性互相独立，那么使用 Bayesian 规则计算后验分布，有：

$$P(Y|S) = \prod_{k=1}^{m} P(y_k|S) = \prod_{i=1}^{n}\prod_{k=1}^{m} P(X_i|S)P(y_k|S, X_i) \quad (5)$$

假设样本集合 $S=\{X_1, X_2, \cdots, X_n\}$ 中 $X_i (1 \leqslant i \leqslant n)$ 服从参数 θ 的分布，密度为 $f(X_i|\theta)$，则对于先验分布 $g(\theta)$，后验密度为：

$$h(\theta|X_1, \cdots, X_n) \approx \prod_{1}^{n} f(X_i|\theta)g(\theta) \quad (6)$$

那么有：

$$P(S|Y) = \int P(S|\theta, y)P(\theta|y)\mathrm{d}\theta \quad (7)$$

所以当 $Y=X_{n+1}$ 时，即在已知 n 个事件的情况下可求第 $n+1$ 个事件发生的概率为：

$$P(X_{n+1}|S)=\int f(X_{n+1}|\theta)h(\theta|S)\mathrm{d}\theta \tag{8}$$

由式(5)、式(7)和式(8)可求出在 X_i 满足二点分布情况下，第 X_{n+1} 个事件发生的概率为：

$$P(X_{n+1}|S)=\prod_{i=1}^{n}P(X_i|S)\int P(X_{n+1}|\theta,X_i)P(\theta|S,X_i)\mathrm{d}\theta$$
$$=\prod_{i=1}^{n}\prod_{k=1}^{m}P(X_i|S)P(x_k|Px_{k+1},\theta,X_i)P(\theta|S,X_i) \tag{9}$$

3.2 将 Bayesian 学习算法应用于本体映射

根据 Bayesian 学习算法的原理，以本体 O_A 和 O_B 为例，用 O_A 和 O_B 中元素 $e_{Ai}\in\{C_A,R_A,I_A\}$ 和 $e_{Bj}\in\{C_B,R_B,I_B\}$ 的语义映射结果构造机器学习的训练集 $S=\{s_1,s_2,\cdots s_k\}$，假设 $s_l(1\leqslant l\leqslant k)$ 之间两两独立，且 O_A 和 O_B 的相似度为 S_{AB}，那么 O_A 和 O_B 间的映射算法如下：

Ontology Mapping Algorithm

输入：$\{C_A,R_A,I_A\}\in O_A$ 和 $\{C_B,R_B,I_B\}\in O_B$，临界参数 t

输出：S_{AB}

步骤 1　初始化训练集，$S=\{s_1,s_2,\cdots s_k\}$

```
l=1;
For (i=1;i≤m;i++)
  {
  e=eAi;//对于本体 OA 中的每一个元素 eAi
    For (j=1;j≤n;j++ )
      if Js(eAi,eBj)>t then sl=Js(eAi,eBj);//在本体 OB 中找到对应元素 eBj
                l++;
  }
k=l;
```

步骤 2　计算 S_{AB}

① 将 $S=\{s_1,s_2,\cdots s_k\}$ 代入式(5)和式(9)计算 s_{k+1} 发生的概率 $P(s_{k+1}|s_1,s_2,\cdots,s_k)$；

② $S_{AB}=\mathrm{argmax}P(s_{k+1}|s_1,s_2,\cdots,s_k)$；

③ 输出 S_{AB}。

其中，基于语义的本体元素映射函数 $Js(e_1,e_2)$ 具体描述如下：

```
Js(e1,e2) //计算分属不同本体的两个元素 e1 和 e2 的相似度
{
  J=0;
  p1=P(e1,e2);
  p2=P(e1,¬e2);
  p3=P(¬e1,e2);
  J=p1/(p1+p2+p3);
}
P(e1,e2) //计算分属不同本体的两个元素 e1 和 e2 相交的概率
{
p=0;
Initialize training set D1={ti|ti∈IA};//初始化训练集 D1 由 OA 中实例组成
Initialize positive training set V1={vi|vi∈e1};//构造属于 e1 的部分正例集合 V1
While (D1≠Φ)
{
  P(ti)=max{P(vi|D1,h)};//找到属于 e1 的新实例 ti
  Add ti to 集合 V1;
  Del ti;
} //在 OA 中找到所有属于 e1 的实例集合 V1
Initialize training set D2={sj|sj∈IB};//初始化训练集 D2 由 OB 中实例组成
Initialize positive training set V2={vj|vj∈e2}∪V1;//构造由 V1 和部分属于 e2 的实例组成的正例集合 V2
While (D2≠Φ)
{
  P(sj)=max{P(vj|D2,h)};
  Add sj to 集合 V2;//找到属于 e1∪e2 的新实例 sj
  Del sj;
  } //在 OB 中找到所有属于 e1∪e2 的实例集合 V2
  p=count(V2)/(count(D1)+count(D2));
}
```

4　实例仿真及对比实验

为了评估 SPIOS 模型的映射准确度和执行效率，采用 H. Do 等人[8]提出的 F-Measure 评估模型，通过对所有映射查全率和精确度的计算来综合评价映射模型的整体性能。

(1)查全率(Recall)：在所有客观存在的映射匹配中正确的映射数目，用 r 表示，其计算公式如下：

$$r=\frac{\text{找出的正确映射数目}}{\text{客观应有的映射数目}} \tag{10}$$

(2)精确度(Precision)：在所有找到的映射匹配中正确的映射数目，用 p 表示，其计算公式如下：

$$p=\frac{\text{找出的正确映射数目}}{\text{所有找出的映射数目}} \tag{11}$$

3)F-Measure 参数：由查全率 r 和精确度 p 构造的综合参数，用 f 表示，其计算公式如下：

$$f=\frac{(b^2+1)pr}{b^2p+r} \tag{12}$$

其中，参数 b 表示查全率 r 和精确度 p 的权重，通常情况下设 $b=1$。

下面根据 F-Measure 评估模型中提出的 3 项参数，分别将使用了 Bayesian 学习算法的 SPIOS 模型与没有使用 Bayesian 学习算法的普通映射模型应用于两种 RBAC 本体的映射中，从而反映 SPIOS 模型的各项性能。实验环境包括 2 台 P2. 0GHz/512M/Cache128k 的 PC，操作系统为 Windows 2003。实验数据来源于文献[9]和文献[10]中提供的两个 RBAC 本体实例，分别称为 RBAC 本体 1 和 RBAC 本体 2，假设其分别应用于两个成员域组成的多域环境下。两个本体的具体描述如表 1 所列。

表 1　关于两个 RBAC 本体的描述

本体名称	总结点数	非叶结点数	深度	总实例数	叶子结点最多实例	一个结点最多子结点
RBAC 本体 1	25	19	3	1220	102	8
RBAC 本体 2	18	11	3	1085	95	5

实验结果如图 2、图 3 和图 4 所示。

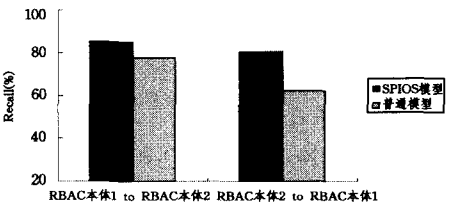

图 2　两种模型关于查全率的比较

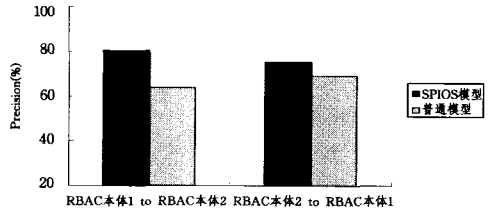

图 3　两种模型关于精确度的比较

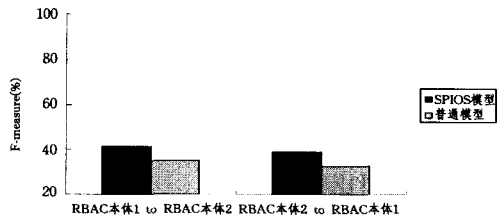

图 4　两种模型关于 F-measure 参数的比较

实验结果表明：①对于相同的映射，使用了 Bayesian 学习算法的 SPIOS 模型较普通模型而言具有更高的查全率和精确度；②SPIOS 模型较普通模型而言，综合性能较好，F-measure 参数较高；③SPIOS 模型较普通模型而言稳定性较好，无论是从 RBAC 本体 1 到 RBAC 本体 2 的映射，还是从 RBAC 本体 2 到 RBAC 本体 1 的映射，查全率、精确度和 F-measure 等各项参数变化幅度不大。

结束语　多域环境下各成员域都使用不同的访问控制体系来维护本域的安全性和自治性，对此不同形式的语法、标记模式和约束为多域安全互操作策略集成提出了挑战，通过引入本体相似度的概念，从语义层构造并映射各成员域的 RBAC 策略；通过 Bayesian 机器学习的算法，从本体的概念、关联和实例等成员中逐步推导本体整体的映射方式，从而将各成员域的访问控制策略映射到全局访问控制本体中，形成适用于多域环境下安全性和自治性的全局 RBAC 策略。

下一步研究的重点将放在本体映射过程的简化和机器学习效率的优化上：通过对本体特征的提取，简化本体间比较和映射过程；通过对机器学习训练集的预处理，提高 Bayesian 学习的效率。

参 考 文 献

[1] Kaushik S, Wijesekera D, Ammann P. Policy-based dissemination of partial web-ontologies[C]// Proceedings of the 2005 Workshop on Secure Web Service. NY, USA, ACM, 2005: 43-52

[2] Bezivin J, Buttner F, Gogolla M, et al. Model Transformations? Transformation Models! [C]// Nierstrasz O, Whittle J, Harel D, et al., eds. Proceedings 9th International Conference on Model Driven Engineering Languages and Systems (MoDELS' 2006). LNCS 4199. Berlin, 2006

[3] Dzbor M, Kubias A, Gridinoc L, et al. The role of access rights in ontology customization[R]. Deliverable 4. 4. 1, NeOn Project, 2007

[4] Kagal L, Berners-Lee T, Connolly D, et al. Self-describing delegation networks for the Web[C]// 7th IEEE Int. Workshop on Policies for Distributed Systems and Networks (POLICY). IEEE, 2006: 205-214

[5] Di W, Jian L, Yabo D, et al. Using semantic Web technologies to specify constraints of RBAC[C]// 6th Int. Conf. on Parallel and Distributed Computing Applications and Technologies (PDCAT). IEEE, 2005: 543-545

[6] 宋炜，张铭. 语义网简明教程[M]. 北京：高等教育出版社，2004

[7] VanRijsbergen. InformationRetrieval[M]. London: Butterworths, 1979

[8] Do H, Melnik S, Rahm E. Comparision of schema matching evaluations[C]// Proceedings of the Seond Int. Workshop on Web Databases (German Informatics Scoiety). 2002

[9] Cirio L, Cruz I F, Tamassia R. A Role and Attribute Based Access Control System Using Semantic Web Technologies[C]// IFIP WG 2. 12 and WG 12. 4 International Workshop on Semantic Web and Web Semantics (SWWS). LNCS 4806. Springer, 2007: 1256-1266

[10] Ionita C M, Osborn S L. Specifying an Access Control Model for Ontologies for the Semantic Web[C]// Jonker W, Petkovic M, eds. SDM2005. LNCS 3674. 2005: 73-85

(上接第 116 页)

[33] Kaddoura Y, King J, Helal A S. Cost-precision tradeoffs in unencumbered floor-based indoor location tracking[C]// Proceedings of the 3rd International Conference on Smart Homes and Health Telematics. 2005: 75-82

[34] Mandal A, Lopes C V, Givargis T, et al. Beep: 3D Indoor Positioning Using Audible Sound[C]// Proceedings of the 2nd Consumer Communications and Networking Conference. 2005: 348-353

[35] Muthukrishnan K, Lijding M, Havinga P. Towards Smart Surroundings: Enabling Techniques and Technologies for Localization [J]. Lecture Notes in Computer Science Location- and Context-Awareness, 2005, 3479: 350-362

[36] Satyanarayanan M. Pervasive Computing: Vision and Challenges [J]. IEEE Personal Communications, 2001, 8(4): 10-17

[37] Hazas M, Ward A. A High Performance Privacy-oriented Location System[C]// Proceedings of the First IEEE International Conference on Pervasive Computing and Communications. 2003: 216-223

[38] Ray S, Ungrangsi R, Pellegrini F D, et al. Robust Location Detection in Emergency Sensor Networks[C]// Proceedings of IEEE INFOCOM. 2003: 1044-1053

[39] Lazos L, Poovendran R, Capkun S. ROPE: robust position estimation in wireless sensor networks[C]// Proceedings of the 4th International Symposium on Information Processing in Sensor Networks. 2005: 324-331

[40] Savarese C, Rabaey J M, Langendoen K. Robust positioning algorithms for distributed ad-hoc wireless sensor networks[C]// Proceedings of the General Track: 2002 USENIX Annual Technical Conference. 2002: 317-327

[41] Gwon Y, Jain R, Kawahara T. Robust Indoor Location Estimation of Stationary and Mobile Users[C]// Proceedings of IEEE INFOCOM. 2004: 1032-1043

[42] Bulusu N, Heidemann J, Estrin D. GPS-less Low-cost Outdoor Location for Very Small Devices [J]. IEEE Personal Communications, 2000, 7(5): 28-34

[43] 唐恩软件科技有限公司. 唐恩 iLocate UWB 定位系统[EB/OL]. http://www.donntech.com/cn/files/rtls/wendang/UWB定位系统概述.pdf, 2009-7

[44] Bouten C V, Koekkoek K T, Verduin M, et al. A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity [J]. IEEE Transactions on Bio-Medical Engineering, 1997, 44(3): 136-47