

基于场景感知的访问控制模型

单棣斌, 杜学绘, 王文娟, 王娜, 刘教迪

(信息工程大学, 河南 郑州 450001)

摘要: 动态访问控制模型是构建大数据动态访问控制系统的理论基础, 而现有访问控制模型大多只能满足单一情景下的动态访问控制, 无法适应大数据上下文环境变化、实体关系变更和客体状态变迁等多类型动态情景中的访问控制。针对上述问题, 在现有访问控制模型的研究的基础上, 对大数据动态因素进行分析, 提出基于场景感知的访问控制 (SAAC, scenario-aware access control) 模型。将各类型动态因素转换为属性、关系等基本元素; 并引入场景信息对各类组成元素进行统一建模; 基于场景信息构建大数据动态访问控制模型, 以实现对多类型动态因素、扩展动态因素的支持。设计 SAAC 模型的工作框架, 并提出框架工作流程对应的基于场景感知的访问控制模型规则学习算法和 SAAC 规则执行算法, 以实现访问控制规则自动学习和动态访问控制决策。通过引入非传递无干扰理论, 分析并验证了对所提模型的安全性。为验证所提模型访问控制策略挖掘方法的有效性, 将 SAAC 模型与 ABAC-L、PBAC-X、DTRM 和 FB-CAAC 等基线模型在 4 个数据集上进行了实验对比。实验结果表明, SAAC 模型及其策略挖掘方法的 ROC 曲线的线下面积、单调性和陡峭度等指标的结果均优于基线模型, 验证了所提模型能够支持多类型动态因素和动态因素扩展, 其挖掘算法所得的访问控制规则的综合质量相对较高。

关键词: 大数据; 访问控制; 动态因素; 场景; 无干扰理论

中图分类号: TP309

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2024003

Scenario-aware access control model

SHAN Dibin, DU Xuehui, WANG Wenjuan, WANG Na, LIU Aodi

Information Engineering University, Zhengzhou 450001, China

Abstract: Dynamic access control model is the theoretical basis for constructing a dynamic access control system for big data. However, most existing access control models can only fulfill dynamic access control in a single scenario and are unable to adapt to access control in multiple types of dynamic scenarios. These scenarios include

收稿日期: 2023-04-25; 修回日期: 2023-08-28

通信作者: 杜学绘, office_paper@163.com

基金项目: 国家自然科学基金 (62102449); 国家重点研发计划 (2018YFB0803603, 2016YFB0501904); 河南省重点研发与推广专项 (222102210069)

Foundation Items: The National Natural Science Foundation of China (62102449), The National Key R&D Program of China (2018YFB0803603, 2016YFB0501904), The Key Research and Development and Promotion Program of Henan Province (222102210069)

引用格式: 单棣斌, 杜学绘, 王文娟, 等. 基于场景感知的访问控制模型[J]. 网络与信息安全学报, 2024, 10(1): 58-78.

Citation Format: SHAN D B, DU X H, WANG W J. Scenario-aware access control model[J]. Chinese Journal of Network and Information Security, 2024, 10(1): 58-78.

changes in the contextual environment of big data, changes in entity relationships, and changes in the state of objects. To address these issues, an analysis was conducted based on the research of existing access control models and the dynamic factors of big data. Subsequently, scenario-aware access control (SAAC) model was proposed, which was based on dynamic factor conversion and scenario unified modeling. All types of dynamic factors were converted into basic elements such as attributes and relationships. Then, scene information was incorporated to model the various types of constituent elements in a unified manner. A big data dynamic access control model was constructed based on scene information to support multi-type dynamic factors and extended dynamic factors. The working framework of the SAAC model was designed, and the SAAC rule learning algorithm and SAAC rule execution algorithm were proposed corresponding to the workflow of the framework. This enabled the automatic learning of access control rules and dynamic access control decision-making. The security of the proposed model was analyzed and verified by introducing the non-transitive non-interference theory. To validate the effectiveness of the access control policy mining method of the proposed model, experimental comparisons were conducted between the SAAC model and baseline models such as ABAC-L, PBAC-X, DTRM, and FB-CAAC using four datasets. The experimental results demonstrate that the SAAC model and its strategy mining method outperforms the baseline models in terms of metrics such as area under the curve AUC, monotonicity, and steepness of the ROC curve. This verification confirms that the proposed model can support multiple types of dynamic factors and dynamic factor extensions, and that the combined effect of the access control rules obtained from its mining algorithm is relatively high.

Keywords: big data, access control, dynamic factors, scenarios, non-interference theory

0 引言

大数据的广泛应用,有力推动了社会进步与经济发展。但是大数据在迅速发展的同时,面临着诸多安全威胁,其中用户越权访问资源是制约大数据资源安全共享、访问受控的主要问题。

大数据资源具有数据量大、动态生成、多源异构、多元关联等特点^[1],而传统访问控制技术采用静态的预先授权模式,难以适应大数据资源的这些特点,无法满足大数据资源动态访问控制的需求。分析大数据的工作机制可以发现,除了实体的属性变化^[2]外,大数据还面临着新的动态访问控制需求^[3-4]。

一是访问大数据资源时环境动态变化^[5],环境不同引发权限不同。例如,在办公地点、工作时间可查阅相关数据资源,而在非办公地点或非工作时间不可访问。这里的地点、时间属于环境条件,也可看作环境的属性信息,这些环境属性信息是动态变化的^[6]。

二是实体关系动态变更,关系变更引发权限变化^[7]。例如,医生可以查阅所属科室收治病人的病历信息,当医生调到其他科室后则不能查阅原科室的病历信息。医生属于科室、病历信息属

于科室等表示医生与科室的关系是访问者(医生)与资源拥有者(科室)之间的关系,这些实体关系信息是动态变更的。

三是客体状态动态变迁,状态转换引发权限改变^[8]。例如,在某科学数据统计过程中,数据资源的原始信息只能被资源的拥有者访问,资源处理的中间环节可以被数据统计方访问,而最后的数据产品可以被普通用户访问。这里依据数据资源不同处理阶段、状态,区分不同访问者各自相应的不同权限,客体的这些不同状态、不同阶段是动态转换的,描述了客体生命周期的不同环节^[9]。

从上述实例可以得出,大数据动态访问控制的本质原因是大数据系统中的某一类或某几类动态因素的变化引发访问者权限随之变更。因此,为适应大数据动态访问控制要求,需要依据这些动态因素进行授权,并随着它们的变化进行权限变更。

访问控制模型是实现访问控制技术、构建访问控制系统的理论基础。同样,大数据动态访问控制的实现需要能够支持动态访问控制的模型作为理论基础。该模型所支持的动态因素决定了其所能支持的动态访问控制类型与能力。访问控制

模型支持动态因素与动态访问控制内容对比如表 1 所示。

现有研究中,大数据所采用的访问控制模型,面临着支持动态因素类型不足、支持潜在动态因素不够的问题,主要表现为。

动态访问控制模型所支持动态因素类别单一或类别较少,不支持多种类型的动态因素。传统模型可以增加对某种动态因素的支持,如 TRBAC^[10] (temporal role-based access control)、GEO-RBAC^[11] (geographical role-based access control)、LoT-RBAC^[12] (location and time-based RBAC) 在 RBAC (role based access control) 模型的基础上增加了时间要素、地点要素,或者二者结合的信息。而现有的动态访问模型,通过对一类动态因素建模,依据此类动态因素进行授权与访问控制,如 ReBAC^[13-14] (relationship based access control) 只能对关系动态因素建立模型,对起源信息、上下文信息都不支持,从而使得模型所能支持动态访问控制内容受到限制,只支持“实体关系”这一类别的动态权限判定^[15]。CAAC^[16-17] (context aware access control)、PBAC^[18] (provenance based access control) 等模型也存在支持动态因素单一等类似问题。

动态访问控制模型对潜在动态因素的发现与扩展不够,不支持对新动态因素的扩展。例如,现有的动态访问控制模型缺少上下文信息的多样化组合^[19]、可推导的实体之间关系^[20]、客体生命周期中的关键状态^[21]等信息,一定程度上限制了模型对应访问控制策略或规则生成的灵活性与自动化。

因此,需要设计一种新型的面向大数据环境的动态访问控制模型,能够融合多种动态因素建模、支持新的动态因素扩展,从而为大数据动态访问控制的实现提供理论支撑。本文的设计思路如下。

1) 针对支持多类型动态因素融合及扩展的模型需求,设计了基于场景的统一建模方法,首先,将动态因素转换为基本元素;其次,利用场景与基本元素的映射关系,实现对多类型动态因素以及扩展的动态因素进行统一表示;最后,建立基于场景感知的访问控制 (SAAC, scenario-aware access control) 模型。

2) 为使模型能够有效地应用于大数据系统动态访问控制,本文设计了基于场景感知的访问控制模型工作框架,框架由访问控制要素、信息库和功能单元等组成,各组成部分通过规则学习、规则执行等环节的执行,实现了场景感知、规则学习和动态决策的动态访问控制过程。

3) 为验证本文所设计访问控制模型的安全性,引入非传递的无干扰理论对基于场景感知的访问控制模型进行了安全性分析。

1 预备知识

1.1 大数据动态因素

大数据^[22]在政府运营、商业、医疗、能源、社交、网络安全和科学研究等领域得到了广泛应用,为各行业、各类别用户提供了丰富的数据服务和知识信息。以医疗大数据平台为例,通过整合个人用户、医院、卫生机构等多方的数据资源,

表 1 访问控制模型支持动态因素与动态访问控制内容对比
Table 1 Compare of access control model supports dynamic factors with dynamic access control content

模型	支持动态授权	动态因素	动态访问控制内容
RBAC	否	无	无
TRBAC	是	时间	时间约束, 随时间动态激活角色
GEO-RBAC	是	地点	结合用户位置授予用户角色
LoT-RBAC	是	时间+地点	结合时间、地点授予用户角色
ABAC	是	属性、环境条件	随属性值的变化、环境条件的变化确定访问者权限
CAAC	是	上下文条件	随时间、地点、客体类别、访问目的等上下文条件的变化判断主体访问权限
ReBAC	是	实体关系	根据主体之间、主客体之间、客体之间的关系判定主体权限
PBAC	是	起源信息	根据起源信息中客体的状态判定主体权限

经过数据生成、收集、处理、分析和应用等工作流程^[23-24]，为医疗行业提供了丰富的、智能化的数据资源、信息产品和知识服务等。

访问控制技术是实现大数据资源安全可控、安全共享的重要保护手段，但大数据资源的数据量大、类型多样、多源汇聚、多方参与、动态处理、应用多元、环境多变等特点，对大数据访问控制策略的制定与应用提出了新的挑战。与大数据访问控制相关的动态因素主要包括以下4个方面。

(1) 属性动态变化

主体、客体、访问操作的属性信息随着访问情景的不同和时间推移而动态发生变化，依据属性信息进行权限判定时必然会发生权限的动态变化。例如，在医疗大数据系统中，某位医生的行政职务由科室医生变为科室主任，那么他相应的访问权限随之增大。基于属性的访问控制（ABAC，attribute based access control）技术^[25]可以通过对属性值的动态获取、与访问控制规则实时匹配，进而动态调整访问控制决策结果。

(2) 上下文环境变化

上下文信息^[26]是用于描述用户、资源或环境等访问控制特定实体状态的相关信息。例如，用户类别、资源类别、时间、位置、访问意图等信息，均被视为可用于访问控制的上下文信息^[27]。

不同的访问会话情景中的时间、地点、执行环境、访问目的等特征信息不同，需要根据上下文变化调整访问权限^[28-29]。例如，医疗数据分析师在工作时间、地点可以对系统中的数据进行分析操作。而在非工作时间、地点只允许浏览分析结果概况信息，不允许执行数据资源分析操作。医生在查阅病历、进修学习等不同任务场景中对大数据平台资源的访问权限不同。因此，需要根据不同的“上下文”信息情景对用户访问资源的行为进行灵活、动态的控制^[30]。

(3) 实体关系变化

多方实体参与的大数据处理过程中，实体之间关系^[31]影响访问者的权限，关系变化带来权限变更^[32]。例如，A医院的医生可以使用大数据治疗方案推荐服务。当A、B两医院建立合作关系后，A医院医生还可以进一步访问治疗方案中B医院案例及患者病例。依据实体关系授权^[33]，当实体满足

关系要求时自动获得相应权限，可简化动态授权。

(4) 客体资源生命周期变化

由于大数据资源是动态生成、转换的，其整个生命周期存在多个阶段，每个阶段有其状态，在不同的状态下，其敏感程度、安全需求是不同的，并随其状态在不断变化。资源在不同阶段、不同状态或形态情况下，访问者的访问权限不同^[34]。例如，医生撰写、阅读原始病历；已归档病历不能修改；医疗数据分析师可以对汇聚的数据集合进行建模计算，但不能查阅病历中原始信息。所以，需要根据数据资源在其生成、计算、呈现等不同阶段的不同状态，进行动态的访问控制，资源的状态信息^[35]是其动态访问控制的依据。

从这4个方面可以得出，大数据动态访问控制的本质原因是大数据系统中的某一类、某些类动态因素的变化引发访问者权限随之变更。因此，为适应大数据动态访问控制要求，需要依据这些动态因素进行授权，依据它们的变化进行授权变更。

1.2 基本概念

为准确描述大数据动态访问控制模型的建模方法，本文给出了大数据动态访问控制及相关动态因素的定义。

定义 1 大数据动态访问控制：在大数据整个生命周期中，自动地感知其各类动态因素的变化过程，并依据其变化情况相应地调整访问控制规则，实时地进行权限裁决与实施控制。

定义 2 属性（At，attribute）：表示主体 S 、客体 O 、访问操作 Op 或环境条件 E 各自的自身特征信息。

属性可记为 $at = \langle atname, atvalue, atmap \rangle \in At$ ，表示一个由属性名称、属性值、属性映射构成的三元组。其中， $atname$ 表示属性名称， $atvalue$ 表示属性值， $Type(atvalue) = \{atomic, set\}$ 表示属性值的类型为原子值型或集合型。 $atmap$ 表示属性名称与属性值之间的映射方式， $atmap \in \{=, \neq, \geq, \leq, >, <, \in\}$ ，默认是“=”，也可以为其他方式，如 $\neq$ 、 $>$ 、 $<$ 、 $\in$ 等。

定义 3 关系（ R ，relationship）：表示主体 S 、客体 O 、访问操作 Op 或环境条件 E 之间的联系。

关系 r 由关系类型、关系值（内容）二元组组成^[20, 36]，记为 $r = \langle rtype, rvalue \rangle$ ，其中 $rtype$ 表

示关系类型, $rvalue$ 表示关系值。

定义 4 环境条件 (E, environments conditions)^[37]: 表示当访问请求时的可检测到的情景信息, 如时间、位置、威胁等级和温度等。

定义 5 上下文信息 (CI, context information): 用于描述用户、资源或环境等访问控制特定实体的状态的相关信息^[30], 如用户类别、资源类别、时间、位置、访问意图等信息, 均被视为可用于访问控制的上下文信息^[38-39]。

为实现对动态因素的统一表示与描述, 本文提出场景及相关概念。

定义 6 场景 (Sc, scenario): 表示属性、关系及其通过计算得到的特征信息。场景包括基本场景、高级场景。场景信息记为

$$Sc = \{sc_1, sc_2, \dots, sc_m\} \quad (1)$$

定义 7 基本场景 (BSc, basic scenario): 表示可以直接获取的主体 S 、客体 O 、访问操作 Op 或环境条件 E 的属性和它们之间关系。根据定义, 基本场景信息包括属性集合 At 、关系集合 R , 表示为

$$BSc = At \cup R \quad (2)$$

定义 8 高级场景 (ASc, advanced scenario):

表示由基本场景通过某种计算方法得到的特征信息。计算方法可以包括组合、推理、派生等方式或者机器学习、深度学习等更高级的运算方法。高级场景信息是由基本场景信息通过某种运算后得到的特征信息, 表示为

$$ASc = \{asc_i | asc_i \in F(BSc)\} \quad (3)$$

其中, F 表示计算高级场景信息的函数方法, 即场景感知方法。

定义 9 场景感知 (SA, scenario-aware): 由基本场景信息学习获得高级场景信息的方法或过程被称为场景感知。

2 SAAC 模型

为实现支持多种类型动态因素、支持动态因素扩展的建模目标, 首先, 需要对各种类型的动态因素、可扩展的新动态因素进行分析, 将其转换为属性、关系等基本的组成元素; 然后, 对各类组成元素进行统一建模, 将其统一表示为场景信息; 最后, 基于场景信息构建动态访问控制模型。大数据动态访问控制模型建模方法如图 1 所示。

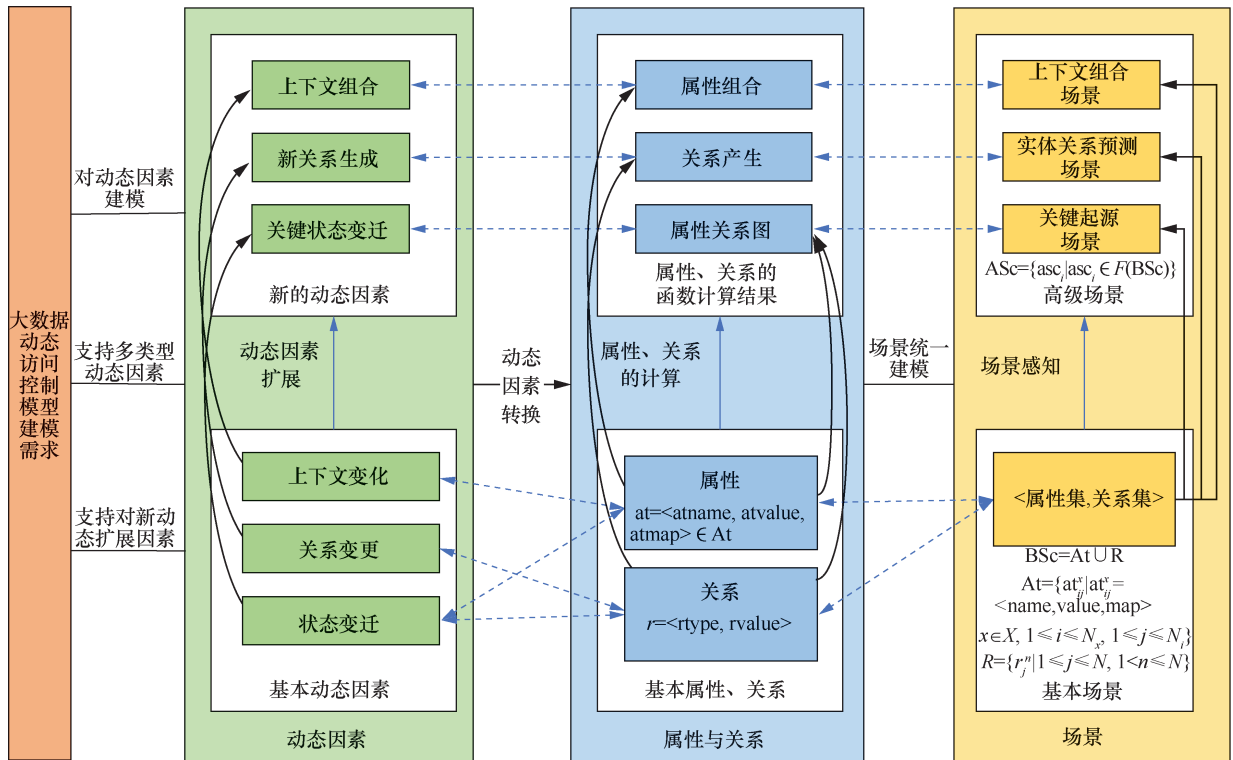


图 1 大数据动态访问控制模型建模方法

Figure 1 Dynamic access control model modeling method for big data

2.1 动态因素转换

通过对大数据特点及大数据动态访问控制分析可知，引发权限变更的动态因素种类比较多，而且会扩展到新的动态因素，难以对其进行统一建模。为解决此问题，本文从动态因素的共性特征入手，将其转换为属性（详见定义2）、关系（详见定义3）等基本的组成元素，为后续统一建模提供支撑。

结合上文所述大数据动态访问控制需求，分析上下文变化、关系变更、状态变迁3种基本动态因素分别和上下文组合、新关系生成、关键状态变迁3种新的动态因素的关系，将其转换为属性、关系等基本元素。

(1) 上下文变化与上下文组合

在ABAC模型中，用于描述访问情景的环境条件（详见定义4）独立于主体或客体，而本文对访问时可检测到的情景信息进行了扩展，不仅包括时间、位置等环境条件，还包括主体类别、客体类别、访问意图等主体、客体的属性信息。为了避免概念混淆，本文采用上下文信息CI（详见定义5）来表示这些内容。

从上下文信息概念可以得出，虽然上下文的种类不同，但它们都代表了用户、资源或环境等自身的特征信息，都可以通过它们的属性来表示，即将上下文信息此类动态因素，转换为主体、客体、访问操作、环境条件的属性信息，记为

$$\begin{aligned} \text{ci}_i^x = \{ \text{at}_{ij}^x \mid x \in X, 1 \leq i \leq N_x, \\ 1 \leq j \leq N_i, \text{ci}_i^x \in \text{CI}^x \subset \text{CI} \} \end{aligned} \quad (4)$$

其中， CI^x 表示类别为 x 的上下文信息集合，它是上下文信息集合CI的子集。 ci_i^x 表示类别为 x 的第 i 种上下文信息集合，包含属性 at_{ij}^x 。

$X = \{S, O, \text{Op}, E\}$ 表示上下文的类别集合，主要包括主体 S 、客体 O 、访问操作 Op 、环境条件 E 共4个类别。每个类别又包括 N_x 种上下文，如环境条件 E 包括时间、地点等不同的上下文，而第 i 种上下文包括 N_i 个属性值， at_{ij}^x 表示类别为 x 的第 i 种上下文的第 j 个属性值。

上下文的变化是指其内容或取值的变化，即对应属性值的变化，记为

$$\begin{aligned} \text{Change}(\text{ci}_i^x) = \{ \text{at}_{ij}^x \cdot \text{atvalue} \rightarrow \text{at}_{ij}^x \cdot \text{atvalue}' \mid x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i \} \end{aligned} \quad (5)$$

在多数访问情景中，不只是单一的上下文，而是多类别上下文的组合。上下文的组合可以由多类别上下文所对应的属性组合构成，记为

$$\begin{aligned} C(\text{CI}) = C(\{ \text{ci}_i^x \mid \text{ci}_i^x \in \text{CI}, x \in X, 1 \leq i \leq N_x \}) = \\ C(\{ \text{at}_{ij}^x \mid \text{at}_{ij}^x \in \text{ci}_i^x, x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i \}) \end{aligned} \quad (6)$$

其中， $C(\text{CI})$ 表示由集合CI中的元素通过组合方式构成的集合信息。

通过上述分析，可将上下文变化（基本动态因素）、上下文组合（扩展动态因素）等通过上下文的属性及其属性值变化来表示。换言之，可以将上下文信息此类动态因素转换为属性信息的变化。

(2) 关系变更与新关系生成

由于大数据多源汇聚、多方参与的特点，访问者的权限与实体之间的关系密切相关，并随着关系的变更而动态变化。大数据资源的生成大多是将多个数据源进行融合汇聚而成的数据资源，数据资源的访问权限与各个数据源的安全需求相关，受到它们之间关系的影响。大数据资源的处理与应用等各个环节都涉及多方实体的参与，如资源拥有者、资源运营方、服务提供方、访问用户等，使得资源的访问权限与各个参与方相关、与它们之间的关系紧密关联。

实体关系表示大数据资源相关实体之间的关系。同类别要素之间存在关系，不同类别要素之间也可以存在关系，如主体-主体关系、客体-客体关系、主体-客体关系、主体-操作关系等。双方参与的关系被称为二元关系，如主体-客体关系；三方参与的关系被称为三元关系，如作者-论文-作者关系，即主体-客体-主体关系，表示属于同一论文的作者，即论文的共同作者。显然，对资源的访问权限与上述实体关系相关，并且随着关系取值的动态变更，访问权限会发生动态变化。

此外，实体关系可以得到扩展。一方面，与关系相关的实体数量可以不断扩展，当 n 个实体之间存在关系，则表示为 n 元关系，称 n 为关系的阶。另一方面，关系相关的要素可以扩展为访问操作、环境条件等，而不仅仅是主体与客体。例如，用户U提交（submit操作）论文F，即主

体-操作-客体关系, 此关系实例表示客体在主体操作后的转化过程。

关系指主体、客体、操作、环境之间相互关联的特征, 如主体与主体关系、客体与客体关系、主体与客体关系、主体与操作关系、操作与客体关系、不同操作之间关系、客体与环境关系等。假设主体 S 、客体 O 、访问操作 Op 或环境条件 E 之间的二元关系 r^2 , 记为

$$r^2 \subseteq x \times x, x \in X \quad (7)$$

n 元关系 r^n , 记为

$$r^n \subseteq \prod_{i=1}^n x_i, x_i \in X, n \in N \quad (8)$$

由于关系由关系类型、关系值组成, 所以关系变更包括关系值变化和关系类型变化两类。其中, 关系值的变化表示为

$$\text{ChangeRV}(r^n) = r^n.rvalue \rightarrow r^n.rvalue' \quad (9)$$

而关系类型的变化相当于建立了新的关系, 可引发关系类型、关系值的变化, 甚至关系阶的变化。关系类型变化表示为

$$\begin{aligned} \text{ChangeRT}(r^n) &= r^n \rightarrow r^m (m, n \in N) = \\ &< r^n.rtype, r^n.rvalue > \rightarrow < r^m.rtype, r^m.rvalue > \end{aligned} \quad (10)$$

依据已有的实体关系集合可以推导出潜在的、目前未知的实体关系, 从而得到可用于访问控制决策的扩展动态因素, 使得访问权限随着新的实体关系的出现与变更而动态变化。根据式(10)可得出, 可由已有关系集合 R 生成某一个新的关系 $r^m (m, n \in N)$, 表示为

$$\begin{aligned} R \rightarrow r^m &= \{r_j^n \mid 1 \leq j \leq N, 1 \leq n \leq N\} \rightarrow \\ &< r^m.rtype, r^m.rvalue > \end{aligned} \quad (11)$$

通过上述分析, 关系变更(基本动态因素)、新关系生成(扩展动态因素)等可通过关系值变化或者关系类型变化来表示。

(3) 状态变迁与关键状态变迁

由于大数据自身特点及其工作机制, 大数据资源经历了动态地生成、传输、合成、运算、应用、销毁等一系列过程。在大数据资源动态的生命周期中, 资源(即客体)呈现了不同的、不断转换的状态和形式, 而资源在不同阶段、状态、形式时, 访问者的访问权限是不同的。例如, 在

大数据汇聚处理过程中, 原始数据资源只可被资源拥有者管理, 数据处理过程中可被数据运营方操作, 而普通用户只能访问数据服务产品。所以, 客体的状态变迁是大数据资源访问控制的动态因素之一。

客体状态变迁的过程一般是由起源数据描述的。起源数据^[8]是对资源信息生命演变过程的重要记录, 描述了数据产生并随时间推移而演变的整个过程, 是实现依据客体状态变化对大数据资源处理过程进行动态控制的关键要素, 一般通过图结构(如 $OPM^{[40]}$ 、 $PROV^{[41]}$)描述客体状态转换的过程。在起源数据图(后简称起源图)中, 主体、客体、操作作为节点, 它们之间的关系作为边。由于起源图中的节点不仅仅有节点类型, 还包括其所对应的具体某一个主体、客体或操作的特征信息, 如客体的名称、内容、长度、关键字、生成时间等属性信息, 所以可以通过节点的属性信息表示起源图的节点。

起源图表示为 $G = (V, E, At, R, f, g)$, V 、 E 分别表示起源图的节点集合和边集合, At 、 R 分别表示对应的属性集合和关系集合, f 和 g 分别表示节点集-属性集、边集-关系集的映射函数。

$$\begin{aligned} f: V &\rightarrow At = V \rightarrow \\ &\{v.at_j^x \mid v \in V, x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i\} \end{aligned} \quad (12)$$

$$\begin{aligned} g: E &\rightarrow R = E \rightarrow \\ &\{< r_e.rtype, r_e.rvalue > \mid r_e \in R, e \in E\} \end{aligned} \quad (13)$$

通过上述映射函数, 起源图的节点、边分别可以用主体、客体和操作的属性以及它们之间的关系进行表示。所以, 起源图可转换为由属性(表示节点)、关系(表示边)组成的图结构, 后文简称为属性关系图。某一个客体状态变迁的过程是起源图的子图, 也就是属性关系图的子图。因此, 状态变迁的过程 G' 可表示为

$$G' = (V', E', At', R', f', g') \subseteq G \quad (14)$$

虽然起源图描述了客体的状态转换过程, 但是拥有海量资源的大数据起源图的规模巨大, 且包含大量与访问控制无关的、冗余的节点与边。因此, 需要从起源图中筛选与动态访问控制相关的关键节点与边, 用于刻画和描述客体的关键状态转换过程。 G' 中的关键状态变迁过程 G'' 可表

示为

$$G'' = (V'', E'', At'', R'', f'', g'') \subseteq G' \subseteq G \quad (15)$$

因此，状态变迁过程及其关键状态变迁的过程，都可以转换为属性、关系这两类基本元素。

综合上述情况，可以得出大数据访问控制的动态因素可以转换为属性、关系两类基本元素。换言之，通过属性、元素以及它们的函数运算结果可以构成这些动态因素，并支持扩展为新的动态因素。

2.2 场景统一建模

2.1 节对各类动态因素进行了分析，并将其转换为属性和关系两类基本元素。通过属性、关系及它们的函数运算结果可以实现对多类别动态因素进行描述，并可实现对新的动态因素的扩展。为了进一步实现对动态因素的统一建模，需要对属性、关系及其函数运算统一进行定义与描述。

2.2.1 基于场景的属性、关系的统一表示

依据场景信息相关定义（详见定义6），属性信息、关系信息属于基本场景信息（详见定义7），可通过场景信息表示，分别记为

$$at_{ij}^x \in \text{BSc.At}, x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i \quad (16)$$

$$r_j^n \in \text{BSc.R}, n \in N, 1 \leq j \leq N \quad (17)$$

对于属性、关系的函数运算结果，如属性组合、关系产生、属性关系图等，都可由场景信息统一表示。

属性组合是由多类别（如主体、客体、操作、环境等）、多种（如主体类别中的各种属性）属性组合在一起构成的，可通过场景信息表示为

$$C(\{at_{ij}^x | x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i\}) \subset C(\text{BSc.At}) \subset \text{ASc.At} \quad (18)$$

关系产生是由已有关系集合 $\{R_j^n | 1 \leq j \leq N, 1 < n \leq N\}$ 生成新的关系信息 r^m ，通过场景信息表示为

$$\{r_j^n | 1 \leq j \leq N, 1 < n \leq N\} \rightarrow \langle r^m.rtype, r^m.rvalue \rangle \in \text{BSc.R} \rightarrow \text{ASc.R} \quad (19)$$

属性关系图 $G = (V, E, At, R, f, g)$ 是由属性（表示节点）、关系（表示边）组成的图结构，通

过场景信息表示图中的节点和边，分别记为

$$f : \{v.at_{ij}^x | v \in V, x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i\} \rightarrow \text{BSc.At} \quad (20)$$

$$g : \{\langle r_e.rtype, r_e.rvalue \rangle | r_e \in R, e \in E\} \rightarrow \text{BSc.R} \quad (21)$$

2.2.2 基于场景的动态因素的统一表示

与大数据访问控制相关的动态因素、扩展得到的新的动态因素可以转换为属性、关系等基本元素，而基本元素属性、关系及其函数运算结果又可以统一用场景表示，所以这些动态因素可以统一由场景信息建模和表示，主要分为上下文组合、实体关系预测和关键起源3类高级场景信息（详见定义8）。

(1) 上下文组合场景

上下文变化可由基本场景的属性集的属性信息变化表示，由式(5)、式(16)可得

$$\text{Change}(ci_i^x) = \{\text{BSc.At.at}_{ij}^x.atvalue \rightarrow \text{BSc.At.at}_{ij}^x.atvalue' | x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i\} \quad (22)$$

作为扩展的动态因素，上下文组合可以通过场景信息的组合进行描述。本文将这种场景信息的组合称为上下文组合场景，它是由基本场景信息通过组合方法得到的高级场景信息。根据式(6)、式(18)可得

$$C(CI) = C(\{at_{ij}^x | at_{ij}^x \in ci_i^x, x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i\}) \subset C(\{at_{ij}^x | x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i\}) \subset C(\text{BSc.At}) \subset \text{ASc.At} \quad (23)$$

其中， $x \in X, 1 \leq k \leq |X|, 1 \leq i \leq N_x, 1 \leq j \leq N_i$ 。

上下文组合场景表示将访问时的实体、操作、环境条件的属性通过组合方式形成一个多种动态因素相结合的情景。例如，类别为 S_1 的用户在 t_1 时刻、在区域 I_1 访问了类型为 O_1 的资源，当前访问场景包括4个属性（主体类别 S_1 、时间 t_1 、位置 I_1 和客体类型 O_1 ），4个属性的组合元组为 $sc_1 = \langle S_1, t_1, I_1, O_1 \rangle$ ，描述了当前访问场景信息。

(2) 实体关系预测场景

关系变更可由基本场景的关系集的关系信息变化表示，由式(9)、式(17)可得到关系值的变化

表示为

$$\text{ChangeRV}(r'') = \text{BSc}.R.r''.\text{rvalue} \rightarrow \text{BSc}.R.r''.\text{rvalue}' \quad (24)$$

由式(9)、式(10)可得, 关系类型变化表示为

$$\text{ChangeRT}(r'') = \text{BSc}.R.r'' \rightarrow \text{ASc}.R.r''^m, m, n \in N \quad (25)$$

新的实体关系是由关系集合通过关系推导函数得到的, 因此可以由场景信息的关系推导函数统一进行表示。本文将这种场景信息的推导称为关系预测场景, 它是由基本场景信息通过某种场景感知方法得到的高级场景信息。

为实现将新的实体关系推导过程转换为关系预测场景, 根据式(11)、式(19)可得

$$R \rightarrow r''^m = \text{BSc}.R \rightarrow \text{ASc}.R.r''^m (m \in N) \quad (26)$$

将关系预测方法定义为 $\text{Prediction}(\cdot)$, 式(26)可表示为

$$\text{ASc}.R.r''^m = \text{Prediction}(\text{BSc}.R) \quad (27)$$

关系预测场景表示通过已有的主体-主体、主体-客体、客体-客体等关系预测出新的实体之间的关系。例如, 主体关系集合 $sR \subseteq S \times S$, 假设预测方法得到主体 u 和 v 之间存在关系 $r(u, v)$, $r(u, v) \notin sR, r(u, v) \in (S \times S - sR)$ 。将关系 $r(u, v)$ 作为新的高级场景信息 sc_2 , 可生成基于场景信息 sc_2 的访问控制规则, 从而实现主体访问客体时的场景满足 sc_2 时对其进行相应的访问控制决策。

(3) 关键起源场景

根据基本元素属性、关系与基本场景信息的映射关系, 状态变迁过程对应的属性关系图的子图可由基本场景的属性集合、关系集合表示为 $G' = (V', E', \text{ASc}.At', \text{ASc}.R', f', g')$ 。

由式(12)、式(13)、式(20)、式(21)可得到起源图中的节点、边与场景信息的映射函数, 分别记为

$$f': V' \rightarrow \text{BSc}.At' \quad (28)$$

$$g': E' \rightarrow \text{BSc}.R' \quad (29)$$

同理, 关键状态变迁(新扩展的动态因素)过程也可由场景信息相应的属性集合和关系集合表示。

本文将起源图中提取到的关键节点与边的集

合称为关键起源场景, 用于对客体关键状态变迁进行描述, 其关键节点、边的映射函数分别表示为

$$f'': V'' \rightarrow \text{ASc}.At'', V'' \subseteq V, \text{ASc}.At'' \subseteq \text{ASc}.At \quad (30)$$

$$g'': E'' \rightarrow \text{ASc}.R'', E'' \subseteq E, \text{ASc}.R'' \subseteq \text{ASc}.R \quad (31)$$

作为起源图的子图, 关键起源场景利用场景信息统一描述了大数据资源关键的状态转换过程, 可进一步地用于依据场景所描述客体状态转换进行动态访问控制。例如, 大数据系统的起源图描述了客体经过主体访问操作发生状态变化的过程, 起源图中客体节点 o_1 和 o_2 通过若干个处理过程(操作 $op_1, op_2, \dots, op_n$), 经过客体 $o_3, o_4, \dots$ 多个状态生成客体 o_n , 处理过程由主体 $s_1, s_2, \dots, s_m$ 触发。假设通过关键节点与边识别方法, 得到关键节点集为 $V'' = \{o_1, o_2, o_3, o_7, o_n, s_1, s_m, op_1, op_n\}$, 关键边集合 E'' 为由 V'' 中节点之间的边组成的集合, V'' 和 E'' 分别映射到高级场景信息 ASc 的属性集合和关系集合, 得到关键起源场景可表示为 $G'' = (V'', E'', \text{ASc}.At'', \text{ASc}.R'', f'', g'') \subseteq G$ 。

通过上述方法, 场景统一描述了各类基本动态因素、新的扩展动态因素的变化过程, 所以, 可将场景信息作为动态授权与访问控制的依据。

2.3 模型形式化定义

通过上述建模方法, 先将与访问控制相关的大数据动态因素转换为属性、关系等基本元素, 再将它们统一由场景信息表示。本文以 ABAC、ReBAC 模型为基础, 引入场景(scenario)概念对动态因素进行统一建模, 设计了基于场景感知的访问控制模型, 为大数据动态访问控制提供支撑。基于场景的访问控制模型如图 2 所示。

(1) 模型要素

1) 主体集 S 表示发起对客体访问操作请求的用户、进程、代理等实体。主体集为 $S = \{s_1, s_2, \dots, s_m\}$, 其中 $s_i \in S, 1 \leq i \leq |S|$ 表示主体集中的一个元素, 即代表某个主体。

2) 客体集 O 表示主体所访问的资源, 如设备、文件、记录、数据表、进程、程序、网络等。客体可以是静态资源, 也可以是动态资源。动态资源的状态、版本会随着资源处理过程而发生相应的变化。客体集合为 $O = \{o_1, o_2, \dots, o_m\}$, 其中

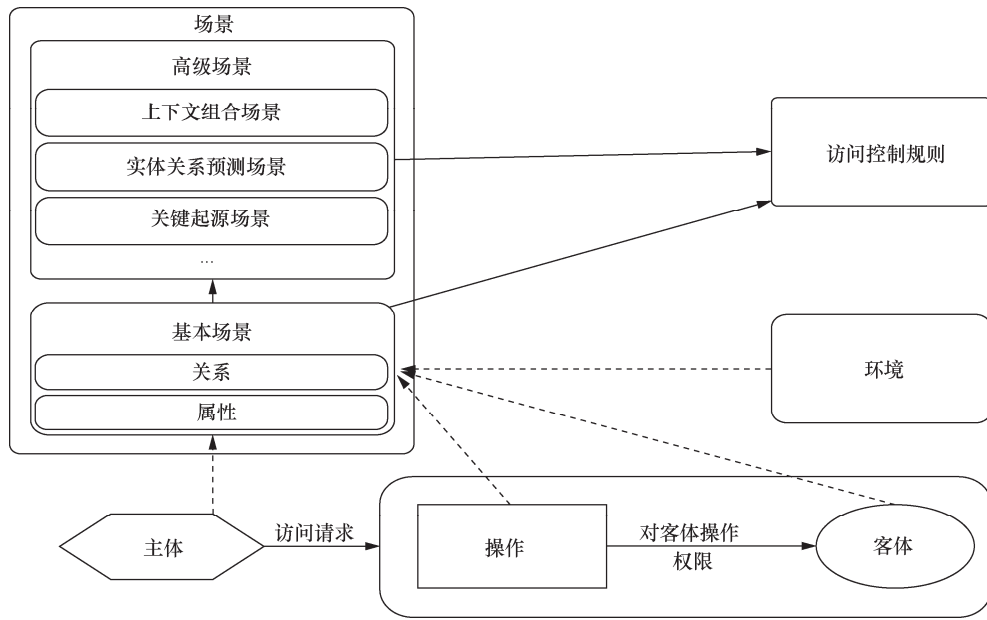


图 2 基于场景的访问控制模型
Figure 2 Scenario-aware access control model

$o_i \in O, 1 \leq i \leq |O|$ 表示客体集中的一个元素, 即代表某个客体。

3) 操作集 Op 表示由主体发起的对客体的访问或执行的功能。访问操作的类型与客体的类型、内容相关, 如读、写、编辑、删除、复制、执行和修改等。操作集合为 $Op = \{op_1, op_2, \dots, op_m\}$, 其中 $op_i \in Op, 1 \leq i \leq |Op|$ 表示操作集合中的一个元素, 即代表某个操作。

4) 环境条件集 E 表示当访问请求时的可检测到的情景信息。环境条件独立于主体或客体, 如时间、位置、威胁等级和温度等。环境条件集合为 $E = \{e_1, e_2, \dots, e_m\}$, 其中 $e_i \in E, 1 \leq i \leq |E|$ 表示环境条件集合中的一个元素, 即代表某个环境条件。

5) 属性集 At 表示主体 S 、客体 O 、访问操作 Op 及环境条件 E 等要素的属性集合。

6) 关系集 R 表示主体 S 、客体 O 、访问操作 Op 、环境条件 E 及属性 At 等之间的关系组成的集合。

7) 场景信息集 Sc 包括基本场景信息 BSc 和高级场景信息 ASc 。基本场景信息包括直接采集获得的属性信息和关系信息。高级场景信息是由基本场景信息通过组合、提取、推理、关联、转换等学习方法得到的更高层次的、综合的特征信

息, 如时间、空间信息的组合, 时间、空间和客体类别的组合, 访问上下文场景, 实体关系的预测与推理, 客体关键状态的变迁过程等场景。

8) 权限集 $Perm$: 表示权限 $\langle op_i, o_j \rangle$ 的集合, 权限 $\langle op_i, o_j \rangle$ 是对客体执行某些操作的许可, 权限集 $Perm$ 是访问操作分配集合 $2^{Op \times O}$ 的子集。 $2^{Op \times O}$ 表示由访问操作到客体的映射关系的集合, 每个访问操作可以与许多客体相关联, 并且每个客体也可以被多个访问操作实施。

9) 访问控制规则集 P 表示一组访问控制规则 p_i 的集合。 p_i 是规则序号 pid , 场景信息 sc_i , 权限信息 $perm_i$ 和访问控制结果 $result$ 的四元组, 表示当主体、客体、操作和环境等满足场景信息时, 主体可以拥有对客体的访问权限为 $perm_i$, 访问控制结果为 $result$ (允许或禁止)。

(2) 映射函数

1) 属性映射函数 MA 表示主体集 S 、客体集 O 、访问操作集 Op 和环境条件集 E 等与属性集 At 之间的映射关系。

2) 关系映射函数 MR 表示为主体集 S 、客体集 O 、访问操作集 Op 、环境条件集 E 和属性集 At 等与关系集合之间的映射关系。

3) 场景权限映射函数 MSP 表示场景集合与权限集的映射关系。

模型形式化定义包括模型要素、映射函数, 如表 2 所示。

2.4 访问控制模型对比

通过上述建模方法, 相比现有的访问控制模型, SAAC 模型对于动态访问控制具有两方面的优势。

1) 支持多种类别基本动态因素。通过属性与关系相结合的基本场景信息, 刻画了上下文条件、属性特征、实体关系等各种动态因素, 使得模型的访问控制策略能够支持类别的基本的动态因素。而 CAAC、ABAC 和 ReBAC 等动态访问控制模型仅支持相对应单一的动态因素, 如 ABAC^[42]模型仅支持属性特征, 而不能支持实体关系等动态因素。

2) 支持动态因素的扩展。由基本场景通过某种计算方法得到的高级场景信息, 可动态表示上下文组合、新实体关系生成、关键状态变迁等新的动态因素, 使得模型的访问控制策略能够支持动态因素扩展。而 RBAC、CAAC、ABAC 和 ReBAC 等访问控制模型所支持的动态因素较为固定, 难以动态变更与扩展。例如, TRBAC^[10]、GEO-RBAC^[11]只能支持在 RBAC 模型的基础上分别增加固定的时间、地点要素, ReBAC^[43]模型本身尚不能支持新实体关系扩展与预测。

3 模型工作框架

3.1 模型框架组成

针对大数据动态访问控制的需求, 本文参考

ABAC 机制的设计思想, 设计了 SAAC 模型的工作框架。基于场景的访问控制模型的工作框架如图 3 所示。

工作框架由 4 个访问控制要素、5 个信息库和 5 个功能单元等组成。其中, 访问控制要素主要包括主体 S (subjects)、客体 O (objects)、操作 Op (operations) 和环境 (E, environments) 这 4 个。访问控制要素已在模型定义中介绍, 此处不再赘述。信息库包括属性信息库 (AIR, attributes information repository)、关系信息库 (RIR, relationships information repository)、访问日志库 (ALR, access log repository)、场景信息库 (SIR, scenario information repository) 和访问控制策略库 (ACPR, access control policy repository) 这 5 个信息库。功能单元包括策略信息点 (PIP, policy information point)、策略挖掘引擎 (PME, policy mining engine)、策略管理点 (PAP, policy administration point)、策略决策点 (PDP, policy decision point) 和策略实施点 (PEP, policy enforcement point) 这 5 个。

(1) 信息库

属性信息库用于存储主体 S、客体 O、访问操作 Op 及环境条件 E 各自的属性信息。

关系信息库用于存储主体 S、客体 O、访问操作 Op、环境条件 E、属性信息 At 之间的关系信息。

访问日志库用于存储大数据系统在运行过程中或测试过程中持续为访问请求生成的审核日志

表 2 SAAC 模型形式化定义
Table 2 Formal definition of SAAC model

模型要素	映射函数
1) 主体集 S: $S = \{s_1, s_2, \dots, s_m\}, s_i \in S, (1 \leq i \leq S)$	属性映射函数 MA:
2) 客体集 O: $O = \{o_1, o_2, \dots, o_m\}, o_i \in O, (1 \leq i \leq O)$	$S \cup O \cup Op \cup E \rightarrow At$
3) 操作集 Op: $Op = \{op_1, op_2, \dots, op_m\}, op_i \in Op, (1 \leq i \leq Op)$	关系映射函数 MR:
4) 环境集 E: $E = \{e_1, e_2, \dots, e_m\}, e_i \in E, (1 \leq i \leq E)$	$S \cup O \cup Op \cup E \cup At \rightarrow R$
5) 属性集 At: $At = \{at_j^x at_j^x = \langle name, value, map \rangle, x \in X, 1 \leq i \leq N_x, 1 \leq j \leq N_i\}$	场景权限映射函数 MSP:
6) 关系集 R: $R = \{r_j^n 1 \leq j \leq N, 1 < n \leq N\}$	$Sc \rightarrow Perm$
7) 场景信息集 Sc: $Sc = BSc \cup ASc, BSc = At \cup R, ASc = F(BSc)$	规则权限映射函数 MPP:
8) 权限集 Perm: $Perm = \{\langle op_i, o_j \rangle op_i \in 2^{Op}, o_j \in 2^{O}\} \subseteq 2^{Op \times O}$	$Policy \rightarrow Perm$
9) 访问控制规则集 P: $P = \{p_i p_i = \langle pid, sc_i, perm_i, result \rangle, pid \in N, sc_i \in Sc, perm_i \in Perm, result \in \{Accept, Deny\}\}$	

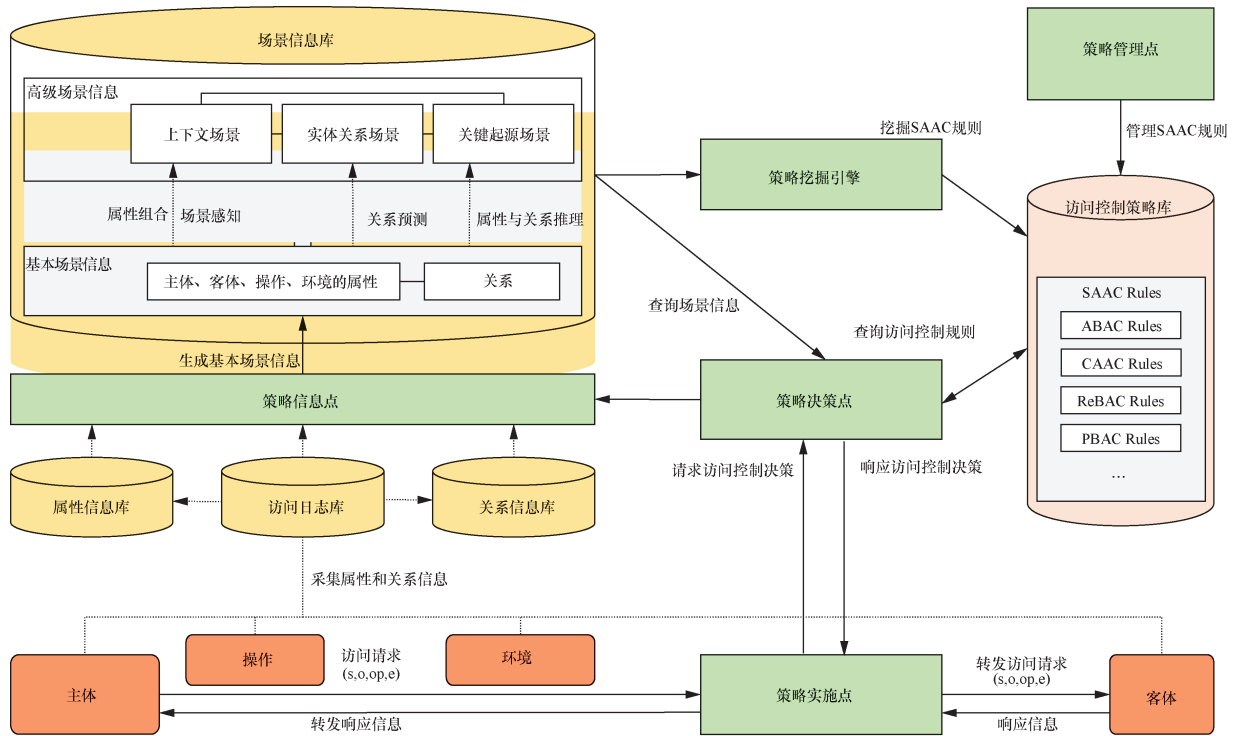


图 3 基于场景的访问控制模型的工作框架
Figure 3 Working framework of the scenario-aware access control model

事件（允许或拒绝）。

场景信息库用于存储基本场景信息 BSc 和高级场景信息 ASc。

访问控制策略库用于存储 SAAC 规则和管理规则。SAAC 规则确定了当满足何种场景信息时，某主体可以对某些客体执行某些操作。依据访问控制规则，对主体请求进行权限判断，判定是否拥有访问目标资源的权限，以及可对资源实施请求的操作类型。SAAC 规则以场景信息、访问日志信息为基础生成，可以支持基于属性、上下文感知、关系、起源等动态因素的访问控制规则，满足大数据动态访问控制的需要。通过管理规则，可实现访问控制规则的存储、验证、更新、优化、冲突发现与消解、执行和失效等管理。

(2) 功能单元

策略信息点负责采集属性和关系信息、生成场景信息（包括直接生成基本场景信息和感知高级场景信息）。采集的数据源包括属性信息库、关系信息库、提出访问请求的主体 S 、访问操作 Op 、访问时的环境 E 、被访问的客体 O ，以及主体访问客体的访问日志信息库。

策略挖掘引擎负责从场景信息库、访问日志库中挖掘信息，以生成基于场景感知的访问控制规则，并将这些规则存储在访问控制策略库中。

策略管理点负责为授权管理员提供一个管理接口，用于查询、修订和测试访问控制规则。

策略决策点负责对 PEP 转发的访问请求进行决策，判断是否允许访问请求通过。决策时依据场景信息库中的场景信息、访问控制策略库中的 SAAC 规则、当前场景信息和访问请求信息进行匹配，得到决策结果。

策略实施点负责依据收到的访问控制决策结果对访问请求与访问响应信息进行控制。对于允许访问的请求或响应信息准予通过，而对于禁止或不确定的请求或响应信息不允许通过。

3.2 框架实施算法

框架的工作流程主要包括 SAAC 规则学习阶段和 SAAC 规则执行阶段。其中，规则学习过程如算法 1 所示。

算法 1 框架实施算法—阶段 1 SAAC 规则学习
输入 AIR, RIR, ALR

输出 SIR, ACPR

```

1) ACPR ← NULL, Rules_c ← NULL
2) ALR_C ← ALR
3) while |ALR_C| > 0 do
4) a_items = Frequent(ALR_C, AIR, ε)
5)   for a_item in a_items do
6)     r_item[] = Index(a_item, RIR)
7)     r_items = r_items + r_item[]
8)   end
9) BSc = a_items + r_items
10) ASc = SA(BSc)
11) Sc = BSc + ASc
12) SIR ← Sc
13)   for Sc_i in Sc do
14)     rule_i = GenerateRules(Sc_i)
15)     Rules_c = Rules_c ∪ rule_i
16)   end
17)   ACPR = ACPR ∪ Rules_c
18)   ALR_C = ALR_C - L(Rules_c)
19) end
20) return SIR, ACPR

```

在算法 1 中, 初始设置 ACPR、学习的访问控制规则集合 Rules_c 为空 (NULL), 将访问日志库的日志集合 ALR 赋值给当前日志信息集 ALR_C (第 1)~2)行)。执行最外层循环结构, 调用 FP-growth 算法^[44]的频繁项函数 Frequent(), 从中获得其频繁项集合 a_items, 这些频繁项集是属性集, ε 是频繁项阈值 (第 3)~4)行)。依次对频繁项集 a_items 中的每一个项 (属性) a_item, 查询 RIR 中的对应关系 r_item, 得到关系集 r_items (第 5)~8)行)。它与属性集 a_items 合成得到基本场景信息集 BSc, 并生成高级场景信息集 ASc (第 9)~12)行)。从场景信息集中依次选择每一个元素 Sc_i, 调用函数 GenerateRules(), 生成对应的访问控制规则 rule_i, 放入 Rules_c (第 13)~16)行)。更新 ACPR, 并从 ALR_C 中删除 Rules_c 对应的所有日志条目。此过程将继续, 直到覆盖所有 ALR_C, 输出 SIR 和 ACPR (第 17)~19)行)。

通过上述算法, 策略信息点从属性信息库、关系信息库、访问日志库中获得与频繁项相关的

属性和关系信息。属性集与关系信息集形成基本场景信息 BSc, 并以此学习高级场景信息 ASc, 存入场景信息库中。由场景信息 Sc 挖掘学习 SAAC 规则, 存入访问控制规则库 ACPR。

SAAC 规则执行过程如算法 2 所示。

算法 2 框架实施算法-阶段 2 SAAC 规则执行

```

输入 Requests, SIR, ACPR
输出 Rules_R
1) Rules_R ← NULL
2) while (r in Requests != NULL) do
3)   at_r = Collect(r)
4)   sc_r = Find(at_r, SIR)
5)   rule_r = Query(sc_r, ACPR)
6)   Rules_R = Rules_R ∪ rule_r
7) end
8) return Rules_R

```

在算法 2 中, 初始设置当前访问请求 Requests 需要执行的访问控制规则集为 Rules_R 为空 (第 1)行)。执行循环结构, 对 Requests 中的每一个访问请求 r, 采集其属性信息 at_r, 依据 at_r 查询 SIR 获得该请求 r 对应的场景信息 sc_r (第 2)~4)行)。依据 sc_r 查询 ACPR 中对应的访问控制规则 rule_r, 并放入 Rules_R 中 (第 5)~7)行)。依次处理所有访问请求, 返回访问控制规则集 Rules_R (第 8)行)。

4 SAAC 模型安全性分析

本文在借鉴 DDTAC-BSS (distributed dynamic trusted access control based on security subsystem) 模型^[45]安全性分析思想的基础上, 通过引入非传递的无干扰理论^[46]对所提 SAAC 模型进行安全验证。

首先, 定义一个基于 SAAC 模型建立的访问控制系统 M, 系统包括成员、函数、关系 3 部分。基于 SAAC 模型的访问控制系统符号及描述如表 3 所示。系统 $M=(ST, D, A, Out)$ 。其中, 根据非传递的无干扰理论中的相关定义, sources、ipurge 计算分别如式(32)、式(33)所示。根据非传递的无干扰理论, 当系统满足式(34)时, 则满足无干扰要求, 系统是安全的。

表 3 基于 SAAC 模型的访问控制系统符号及描述
Table 3 Notation and description of access control system based on SAAC model

类别	符号	描述
成员	ST	系统状态集, $ST=\{st_0, st_1, \dots, st_n\}$, st_0 表示系统初始状态
	D	安全域, 包括请求域 d_s 、响应域 d_o 、访问控制域 d_a (d_a 可细分为访问控制决策域 d_d 、访问控制实施域 d_e)
	A	行为集, $A=\{a_0, a_1, \dots, a_n\}$, a_i 表示第 i 个行为。根据 SAAC 模型, 行为包括访问请求 $a_{request}$ 、访问响应 $a_{response}$ 、访问裁决 $a_{decision}$ 、访问控制实施 $a_{enforcement}$ 、场景感知 $a_{scaware}$ 、访问控制规则生成 a_{rulesg} 6 类行为。
	Out	输出结果集
函数	$dom: A \rightarrow D$	行为与安全域的映射函数, 返回值是行为所属的安全域。 $dom(a_i) \in D$ 表示行为 a_i 对应的安全域
	$step: ST \times A \rightarrow ST$	系统状态 st_i 在执行行为 a_i 后进入系统状态 st_{i+1} ($0 \leq i \leq n$), 记为 $st_{i+1} = step(st_i, a_i)$
	$run: ST \times A^* \rightarrow ST$	系统状态 st_i 在执行一系列行为 α 后进入系统状态 st_{i+j} ($0 \leq i \leq n$), 记为 $st_{i+j} = run(st_i, \alpha)$ 。其中, α 是行为序列, 表示为 $\alpha = a_{i+1} \circ a_{i+2} \circ \dots \circ a_j$, $\circ$ 表示行为连接
	$output: ST \times A^* \rightarrow Out$	系统状态 st_i 在执行一系列行为 α 后的输出结果, 记为 $out_{i+j} = output(st_i, \alpha)$ 。当 α 只有一次行为时, $output$ 表示单步执行后的输出结果
	$ipurge: A^* \times D \rightarrow A^*$	非传递的消除函数, 从行为序列 α 中消除与给定安全域 d 无干扰的行为, 记为 $ipurge(\alpha, d)$
	$sources: A^* \times D \rightarrow 2^D$	在一个行为序列中识别出那些不应该被删除的行为, 记为 $sources(\alpha, d)$
关系	$\sim >$	干扰关系: 表示对不同安全域之间的信息流的权限
	$\not\sim >$	无干扰关系: 表示禁止不同安全域之间的信息流
	$st_i \sim^d st_j$	观察等价关系: 表示从安全域 d 的角度来看, 系统状态 st_i 和状态 st_j 是等价的。 $st_i \sim^d st_j \Rightarrow output(st_i, a) = output(st_j, a)$
	$st_i \approx^c st_j$	子域等价关系: 表示在安全域集的一个子域中, 系统状态 st_i 和状态 st_j 是等价的。 $st_i \approx^c st_j =_{def} (\forall d \in C: st_i \sim^d st_j)$
	$C: st_i \sim^d st_j$	

$$sources(\emptyset, d) = \{d\}$$

$$sources(a \circ \alpha, d) = \begin{cases} sources(\alpha, d) \cup \{dom(a)\}, \text{其中}, \exists v: v \in sources(\alpha, d) \wedge dom(a) \sim > v \\ sources(\alpha, d), \text{其他} \end{cases} \quad (32)$$

$$ipurge(\emptyset, d) = \emptyset \quad ipurge(a \circ \alpha, d) = \begin{cases} a \circ ipurge(\alpha, d), \text{其中}, dom(a) \in sources(a \circ \alpha, d) \\ ipurge(\alpha, d), \text{其他} \end{cases} \quad (33)$$

$$output(run(st_0, \alpha), a) = output(run(st_0, ipurge(\alpha, dom(a))), a) \quad (34)$$

假定系统初始状态 st_0 是安全的, 式(34)等式左侧表示系统从初始状态 st_0 执行行为序列 α 后执行行为 a 的输出结果, 式(34)等式右侧表示行为序列 α 消除对行为 a 所属安全域无干扰的行为, 系统从初始状态 st_0 执行经过消除操作的行为序列后执行行为 a 的输出结果。对基于 SAAC 模型的访问控制系统中, 表示执行访问控制行为时, 请求域的行为不会影响到响应域观察到的系统输出信息, 则请求域不会干扰响应域。为验证 SAAC

系统满足式(34), 引入非传递策略展开定理, 验证系统的安全性。

定理 1 非传递策略展开定理^[46]。 $\sim >$ 满足非传递策略, 系统 M 划分安全域视图, 当 M 满足下列条件时, 对于 $\sim >$ 是安全的。

1) 输出一致性。

$$st_i \sim^{dom(a)} st_j \Rightarrow output(st_i, a) = output(st_j, a) \quad (35)$$

2) 弱单步一致性。

$$\begin{aligned} & \text{st}_i \stackrel{d}{\sim} \text{st}_j \wedge \text{st}_i \stackrel{\text{dom}(a)}{\sim} \text{st}_j \Rightarrow \\ & \text{step}(\text{st}_i, a) \stackrel{d}{\sim} \text{step}(\text{st}_j, a) \end{aligned} \quad (36)$$

3) 局部干扰性。

$$\text{dom}(a) \not\sim d \Rightarrow \text{st}_i \stackrel{d}{\sim} \text{step}(\text{st}_i, a) \quad (37)$$

证明过程参见文献[46]的定理 7 证明, 此处不再赘述。

定理 2 基于 SAAC 模型的访问控制系统, 当请求域执行请求行为时, 访问控制域执行相关行为序列后, 访问控制域与响应域观察到的系统状态是相同的, 即请求域对响应域是非传递无干扰的。

证明

1) 满足非传递策略, 系统 M 划分安全域视图。

基于 SAAC 模型的访问控制系统, 划分了请求域 d_s 、响应域 d_o 、访问控制域 d_a , 由访问控制域将请求域和响应域分割, 满足 $(d_s \sim d_a) \wedge (d_a \sim d_o) \wedge (d_s \not\sim d_o)$, 即请求域与访问控制域相关, 访问控制域与响应域相关, 但请求域对响应域在关系 $\sim$ 下是不直接相关的。

2) 满足输出一致性。

$$\text{st}_i \stackrel{\text{dom}(a)}{\sim} \text{st}_j \Rightarrow \text{output}(\text{st}_i, a) = \text{output}(\text{st}_j, a) \text{ 成立}$$

3) 满足弱单步一致性。

$$\text{st}_i \stackrel{d}{\sim} \text{st}_j \wedge \text{st}_i \stackrel{\text{dom}(a)}{\sim} \text{st}_j \Rightarrow \text{step}(\text{st}_i, a) \stackrel{d}{\sim} \text{step}(\text{st}_j, a)$$

成立。

4) 满足局部干扰性。

需要证明子域等价关系满足

$$\text{st}_i \stackrel{\text{sources}(a, d)}{\approx} \text{st}_j \Rightarrow \text{run}(\text{st}_i, a) \stackrel{d}{\sim} \text{run}(\text{st}_j, \text{ipurge}(a, d)) \quad (38)$$

通过归纳法进行证明。

首先, 令 $a = \emptyset$, 式(38)成立。

其次, 假设行为序列 a 的长度为 n 时式(38)成立, 如果在长度为 $n+1$ 时也成立, 即执行行为序列 $a \circ \alpha$ 后也成立, 则满足局部干扰性。所以, 如果式(38)成立, 则需要证明

$$\begin{aligned} & \text{st}_i \stackrel{\text{sources}(a \circ \alpha, d)}{\approx} \text{st}_j \Rightarrow \text{run}(\text{st}_i, a \circ \alpha) \\ & \stackrel{d}{\sim} \text{run}(\text{st}_j, \text{ipurge}(a \circ \alpha, d)) \end{aligned} \quad (39)$$

① 当 $\text{dom}(a) \in \text{sources}(a \circ \alpha, d)$ 时, 由式(33)可得

$$\text{ipurge}(a \circ \alpha, d) = a \circ \text{ipurge}(\alpha, d) \quad (40)$$

此时, 根据式(38)、式(40), 对式(39)进行推导可得

$$\begin{aligned} & \text{st}_i \stackrel{\text{sources}(a \circ \alpha, d)}{\approx} \text{st}_j \Rightarrow \\ & \text{step}(\text{st}_i, a) \stackrel{\text{sources}(\alpha, d)}{\approx} \text{step}(\text{st}_j, a) \Rightarrow \\ & \text{run}(\text{step}(\text{st}_i, a), \alpha) \stackrel{d}{\sim} \text{run}(\text{step}(\text{st}_j, a), \text{ipurge}(\alpha, d)) \Rightarrow \\ & \text{run}(\text{st}_i, a \circ \alpha) \stackrel{d}{\sim} \text{run}(\text{st}_j, a \circ \text{ipurge}(\alpha, d)) \Rightarrow \\ & \text{run}(\text{st}_i, a \circ \alpha) \stackrel{d}{\sim} \text{run}(\text{st}_j, \text{ipurge}(a \circ \alpha, d)) \end{aligned} \quad (41)$$

由式(41)可得式(39)成立。

② 当 $\text{dom}(a) \notin \text{sources}(a \circ \alpha, d)$ 时, 由式(33)可得

$$\text{ipurge}(a \circ \alpha, d) = \text{ipurge}(\alpha, d) \quad (42)$$

$$\text{dom}(a) \notin \text{sources}(a \circ \alpha, d) \Rightarrow \text{st}_i \stackrel{\text{sources}(\alpha, d)}{\approx} \text{step}(\text{st}_i, a) \quad (43)$$

因为 $\text{sources}(\alpha, d) \subseteq \text{sources}(a \circ \alpha, d)$,

$\text{st}_i \stackrel{\text{sources}(a \circ \alpha, d)}{\approx} \text{st}_j$, 所以

$$\text{st}_i \stackrel{\text{sources}(\alpha, d)}{\approx} \text{st}_j \quad (44)$$

$$\text{step}(\text{st}_i, a) \stackrel{\text{sources}(\alpha, d)}{\approx} \text{st}_j \quad (45)$$

此时, 根据式(38)、式(42)、式(45), 对式(39)进行推导可得

$$\begin{aligned} & \text{st}_i \stackrel{\text{sources}(a \circ \alpha, d)}{\approx} \text{st}_j \\ & \Rightarrow \text{step}(\text{st}_i, a) \stackrel{\text{sources}(\alpha, d)}{\approx} \text{st}_j \\ & \Rightarrow \text{run}(\text{step}(\text{st}_i, a), \alpha) \stackrel{d}{\sim} \text{run}(\text{st}_j, \text{ipurge}(\alpha, d)) \\ & \Rightarrow \text{run}(\text{st}_i, a \circ \alpha) \stackrel{d}{\sim} \text{run}(\text{st}_j, \text{ipurge}(a \circ \alpha, d)) \end{aligned} \quad (46)$$

由式(46)可得式(39)成立。

综合①②这两种情况可得, 当序列长度为 $n+1$ 时, 即执行行为序列 $a \circ \alpha$ 后, 式(39)成立, 归纳推导结束。定理 2 证毕。

令 a 为访问控制域的行为序列, 令 a 为请求

域请求行为 $a_{request}$ ，通过上述定理 2 的证明，可以得到，请求域 d_s 与访问控制域 d_a 相关联，访问控制域 d_a 与响应域 d_o 相关联，但请求域 d_s 对响应域 d_o 是非传递无干扰的。同理，当 a 为场景感知的行为 $a_{scaware}$ 、基于场景感知的访问控制规则生成 a_{rulesg} 等访问控制域的行为时，没有影响请求域 d_s 对响应域 d_o 的非传递无干扰性。所以，基于 SAAC 模型的访问控制系统 M 中的请求域 d_s 对响应域 d_o 是安全的。

5 实验结果与数据分析

5.1 数据集

为验证访问控制规则挖掘学习效果，本文选用 University^[14]、ACM^[47]、Healthcare^[14] 和 Appusage^[48] 这 4 个数据集作为规则匹配的基准测试数据。这些数据集在 ABAC、ReBAC、PBAC 和 CAAC 等访问控制模型的策略挖掘实验中得到了广泛应用，可用于本文所提 SAAC 模型访问控制规则挖掘、与基线访问控制模型对比等实验分析。上述每个数据集划分为测试数据和验证数据两部分，占比分别为 80%、20%。测试数据用于挖掘访问控制规则，验证数据用于验证挖掘规则的质量。

5.2 基线模型与评估指标

为实现对本文所提 SAAC 模型的访问控制规则挖掘学习结果进行对比分析，选用 ABAC-L^[49]、PBAC-X^[21]、DTRM^[14] 和 FB-CAAC^[27] 这 4 个访问控制模型的规则挖掘方法作为基线模型。

本文将真阳性率 (TPR, true positive rate)、假阳性率 (FPR, false positive rate) 分别作为衡量通过挖掘算法所得访问控制规则的权限覆盖率^[50]、过度授权比率^[50]的指标，并通过接收者操作特性 (ROC, receiver operating characteristic) 曲线综合判定挖掘所得访问控制规则的质量^[50]。

(1) TPR

本文将 TPR 作为衡量权限覆盖率的指标，表示在验证数据集允许访问的事件中，挖掘所得规则允许访问事件数量所占的比例。TPR 可用于衡量挖掘规则集合授权是否不足，即作为评估整个规则集的权限覆盖率的指标。TPR 计算方法为

$$TPR = TP / (TP + FN) \quad (47)$$

对于验证数据集允许访问的事件，当挖掘所

得规则都允许访问时，则 $FN=0$ ， $TPR=1$ ，表示规则授权充分，没有缺少授权。当验证数据集允许访问而挖掘所得规则禁止访问的事件数量增加时，TPR 随之降低，表示规则授权不足，其极限值为 0。

(2) FPR

本文将 FPR 作为衡量过度授权比率的指标，表示在验证数据集禁止访问的事件中，而挖掘所得规则却允许访问事件数量所占的比例。FPR 可用于衡量挖掘规则集合授权是否超出应有权限，即作为评估整个规则集的过度授权比率的指标。FPR 计算方法为

$$FPR = FP / (FP + TN) \quad (48)$$

对于验证数据集禁止访问的事件，当挖掘所得规则都禁止访问时，则 $FP=0$ ， $FPR=0$ ，表示规则没有过度授权。当验证数据集禁止访问而挖掘所得规则允许访问的事件数量增加时，FPR 随之增加，表示规则过度授权，其极限值为 1。

(3) ROC 曲线

ROC 曲线描述了 TPR (纵坐标) 和 FPR (横坐标) 之间的关系，其衡量指标包括曲线下面积 (AUC, area under curve)、曲线的单调性和陡峭程度。

ROC 曲线下面积表示预测的正例排在负例前面的概率。所以 AUC 分数越高，表示所得访问控制规则的质量越好。

ROC 曲线的单调性反映了 TPR 值随 FPR 增加的变化趋势。理想状态下，TPR 值应随着 FPR 的增加而单调增加。所以 ROC 曲线的单调性越好，则它的平滑度越好，表示挖掘所得访问控制规则的权限覆盖率随着过度授权比率增加而增长。

ROC 曲线的陡峭程度反映了 TPR 值随 FPR 增加的变化速度。理想状态下，TPR 值应随着 FPR 的增加而迅速增加，达到或接近峰值。ROC 曲线的陡峭程度越大，则越有降低 FPR、升高 TPR 的趋势。它表示挖掘所得的访问控制规则，既可保持较低的过度授权比率，又可获得较高的权限覆盖率。

5.3 实验数据分析

本文首先通过实验分析了频繁项阈值参数与 ROC 曲线的关系，确定了该参数的较优值；其次，

利用 ROC 曲线对比了 SAAC 访问控制规则挖掘方法与基线模型的访问控制规则挖掘结果,从而综合判定挖掘所得的 SAAC 访问控制规则的质量。

5.3.1 频繁项阈值 ε 与 ROC 曲线关系

频繁项阈值 ε 是 SAAC 访问控制规则挖掘算法调用 FP-growth 算法^[44]时的参数(详见算法 1 第 4)行),用于指定如果模式出现在检查条目的频率不小于 ε ,则该模式被视为“频繁”模式。频繁项阈值 ε 的取值影响了访问控制规则挖掘算法的迭代次数和目标规则的选择,因此是影响 ROC 曲线的关键要素。当 ε 在 $[0.05, 0.1, 0.15, 0.2, 0.25]$ 时,通过 SAAC 访问控制规则挖掘算法在 4 个数据集执行所得 ROC 曲线变化,频繁项阈值 ε 与 ROC 曲线关系如图 4 所示。

实验结果显示, ε 越小,ROC 曲线下面积 AUC 越大。但是,当 $\varepsilon \leq 0.1$ 时,AUC 随 ε 的减少而增长的数量并不明显,且 ROC 曲线出现了波动。例如, $\varepsilon=0.05$ 时,各个数据集的 ROC 曲线下面积 AUC 与 $\varepsilon=0.1$ 时各自的 AUC 值接近,而且数据

集 University、ACM 和 Appusage 的 ROC 曲线($\varepsilon=0.05$)都出现了轻微波动,即 ROC 曲线的单调性不佳,表明其 TPR 值没有完全随着 FPR 值的增加而单调增加。这是由于当 ε 较小时,FP-growth 算法识别的候选项集合增大,会产生更多的候选项进行评估,通常也会得到较高的 AUC 值。但是当 ε 过小时,造成其候选项集合迅速增大,降低了候选项对权限的覆盖程度,从而减缓了其 AUC 的增长速度,而且会由于候选项集合的增大而大幅增加规则挖掘的时间开销。当 $\varepsilon > 0.1$ 时,AUC 随 ε 的增加而迅速减少。这是由于 ε 值较大,FP-growth 算法识别的候选项越来越少,从而 FPR 下降。因此,在当前数据集实验中,选择 0.1 作为 ε 的较优值,既可获得较高的 AUC 值,又可控制规则挖掘的时间开销。

5.3.2 ROC 曲线对比

通过 ROC 曲线可以比较 SAAC 访问控制规则挖掘模型与基线模型挖掘所得规则的质量。SAAC 模型与基线模型(ABAC-L、

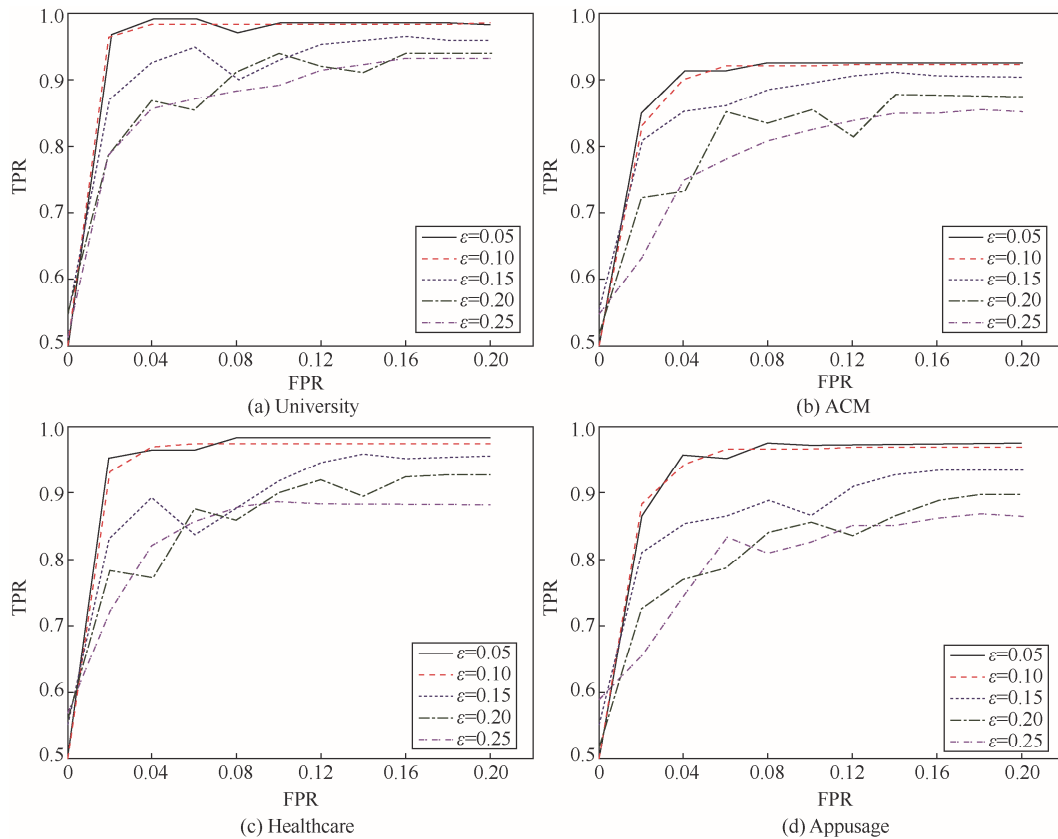


图 4 频繁项阈值 ε 与 ROC 曲线关系
Figure 4 Frequent term threshold ε and ROC curve

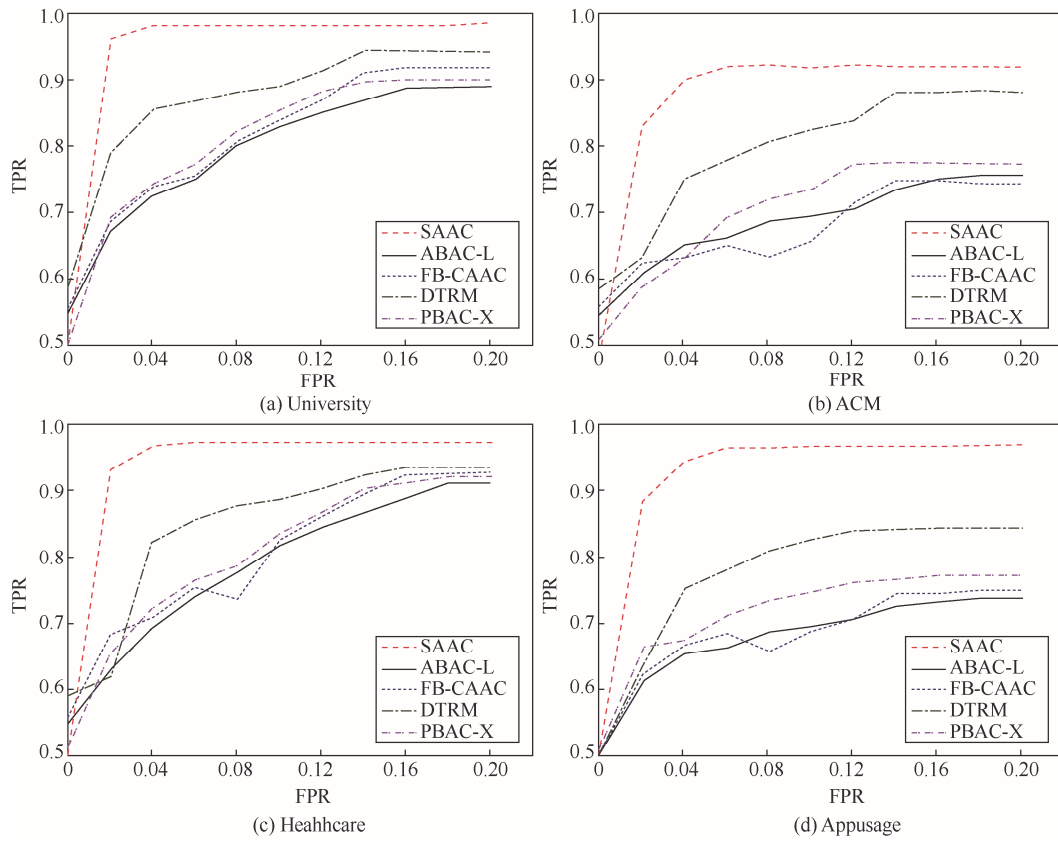


图 5 SAAC 与基线模型的 ROC 曲线比较结果
Figure 5 Results of ROC curve comparison between SAAC and baseline model

PBAC-X、DTRM 和 FB-CAAC 这 4 个)的 ROC 曲线比较结果如图 5 所示。其中, 参数频繁项阈值 $\varepsilon=0.1$ 。

(1) ROC 曲线下面积 AUC

SAAC 对应的 ROC 曲线下面积 AUC 值最大, DTRM 的 AUC 值次之, PBAC-X、FB-CAAC 和 ABAC-L 的 AUC 值相对较小。实验结果表明: 相较于基线模型, SAAC 挖掘所得的访问控制规则集的质量最好。

(2) ROC 曲线单调性

SAAC 的 ROC 曲线单调性最好, 其 TPR 随着 FPR 的增加而单调增加。DTRM、PBAC-X、ABAC-L 的 ROC 曲线随着 FPR 的增长, TPR 的增长率出现了轻微的波动, 如在数据集 ACM 中, DTRM、PBAC-X、ABAC-L 也出现增长率下降后又上升的起伏过程。而 FB-CAAC 的 ROC 曲线出现了随着 FPR 的增长, TPR 反而下降的情况。实验结果表明, SAAC 挖掘的规则 TPR (即规则的权限覆盖率) 的变化趋势优于基线模型。

(3) ROC 曲线陡峭程度

SAAC 的 ROC 曲线的陡峭程度最好, 而 DTRM、PBAC-X、FB-CAAC 和 ABAC-L 等基线模型的 ROC 曲线陡峭程度相对较为平缓。实验结果表明, SAAC 挖掘的规则可以在较小的假阳性率时, 获得较高真阳性率, 既可保持较低过度授权比率, 又可获得较高的权限覆盖率。

上述实验结果表明, 相较于 4 个基线模型, SAAC 挖掘所得的访问控制规则综合效果最优。这是由于 SAAC 访问控制规则挖掘方法在提高规则权限覆盖率、降低过度授权比率和规则复杂度方面具有优势。首先, SAAC 规则包括基于基本场景信息和高级场景信息的访问控制规则, 规则根据实体属性、关系以及它们的动态变化信息等要素进行授权与控制, 相比单纯的属性、关系等信息, 访问控制粒度更加灵活、权限支持动态变化, 使得权限覆盖率得到了提高, 而过度授权比率得到控制。其次, SAAC 规则挖掘方法综合权限覆盖率、过度授权比率

作为评估指标,通过迭代学习后,将权限覆盖率高、过度授权比率小,且复杂度较低的规则放入候选规则集,从而提高了挖掘所得 SAAC 规则的质量。

6 结束语

针对支持多类型动态因素和扩展动态因素的访问控制模型建模需求,本文提出了动态因素转换、场景统一表示的建模方法,建立了基于场景的访问控制模型,构建了模型的工作框架,通过非传递无干扰理论验证了所设计访问控制模型的安全性,通过实验对比验证了 SAAC 规则挖掘方法获得了较好的效果。下一步将深化研究场景信息学习方法,实现上下文属性组合、关系预测等高级场景信息的自动学习,以提升 SAAC 模型在大数据资源平台中的场景感知能力,满足其动态访问控制的需求。

参考文献:

- [1] 李昊, 张敏, 冯登国, 等. 大数据访问控制研究[J]. 计算机学报, 2017, (1): 72-91.
LI H, ZHANG M, FENG D G, et al. Research on access control for big data [J]. Journal of Computer Science, 2017, (1): 72-91.
- [2] SERVOS D, OSBORN S L. Current research and open problems in attribute-based access control[J]. ACM Computing Surveys, 2017, 49(4): 1-45.
- [3] CHEN X, GAO Y, TANG H, et al. Research progress on big data security technology[J]. Scientia Sinica Informationis, 2020, 50(1): 25-66.
- [4] 高振升, 曹利峰, 杜学绘. 基于区块链的访问控制技术研究进展[J]. 网络与信息安全学报, 2021, 7(6): 68-87.
GAO Z S, CAO L F, DU X H. Research progress of access control based on blockchain[J]. Chinese Journal of Network and Information Security, 2021, 7(6): 68-87.
- [5] KAYES A S M, KALARIA R, SARKER I H, et al. A survey of context-aware access control mechanisms for cloud and fog networks: taxonomy and open research issues[J]. Sensors (Basel), 2020, 20(9): 1-34.
- [6] 刘敖迪, 杜学绘, 王娜, 等. 基于深度学习的 ABAC 访问控制策略自动化生成技术[J]. 通信学报, 2020, 41(12): 8-20.
LIU A D, DU X H, WANG N, et al. Automatic Generation technology of ABAC access control policy based on deep learning [J]. Journal on Communications, 2020, 41(12): 8-20.
- [7] 单棣斌, 杜学绘, 王文娟, 等. 基于 GNN 双源学习的访问控制关系预测方法[J]. 网络与信息安全学报, 2022, 8(5): 40-55.
SHAN D B, DU X H, WANG W J, et al. Access control relationship prediction method based on GNN dual source learning[J]. Chinese Journal of Network and Information Security, 2022, 8(5): 40-55.
- [8] PARK J, NGUYEN D, SANDHU R. A provenance-based access control model[C]//Proceedings of 2012 Tenth Annual International Conference on Privacy, Security and Trust (PST). 2012: 137-144.
- [9] AKAICHI I, KIRrane S. Usage control specification, enforcement, and robustness: a survey[J] 2022, arXiv: 2203. 04800[04.23 2023].
- [10] BERTINO E, BONATTI P A, FERRARI E. TRBAC: A temporal role-based access control model[C]//Proceedings of TISSEC2001. 2001: 191-233.
- [11] BERTINO E, CATANIA B, DAMIANI M L, et al. GEO-RBAC: a spatially aware RBAC[C]//Proceedings of the 10th Symposium on Access Control Models and Technologies. 2005: 29-37.
- [12] CHANDRAN S M, JOSHI J B. LoT-RBAC: A location and time-based RBAC model[C]//Proceedings of the International Conference on Web Information Systems Engineering. 2005: 361-375.
- [13] ASIM Y, MALIK A K. A survey on access control techniques for social networks[M]. Information Diffusion Management and Knowledge Sharing. 2020: 319-342.
- [14] BUI T, STOLLER S D. A decision tree learning approach for mining relationship-based access control policies[C]//Proceedings of the 25th ACM Symposium on Access Control Models and Technologies. 2020: 167-178.
- [15] BUI T, STOLLER S D, LE H. Efficient and extensible policy mining for relationship-based access control[C]//Proceedings of the 24th ACM Symposium on Access Control Models and Technologies. 2019: 161-172.
- [16] KAYES A S M, RAHAYU W, DILLON T, et al. Context-aware access control with imprecise context characterization for cloud-based data resources[J]. Futur Gener Comp Syst, 2019, 93: 237-255.
- [17] KAYES A S M, HAN J, RAHAYU W, et al. A policy model and framework for context-aware access control to information resources[J]. Comput J, 2019, 62(5): 670-705.
- [18] BERTOLISSI C, HARTOG J D, ZANNONE N. Using provenance for secure data fusion in cooperative systems[C]//Proceedings of the 24th ACM Symposium on Access Control Models and Technologies. 2019: 185-194.
- [19] YU Y, XIA T, WANG H, et al. Semantic-aware spatio-temporal app usage representation via graph convolutional network[J]. Proc ACM Interact Mob Wearable Ubiquitous Technol, 2020, 4(3): 101:101-124.
- [20] CHAKRABORTY S, SANDHU R. Formal analysis of ReBAC

- policy mining feasibility[C]//Proceedings of CODASPY '21. 2021: 197-207.
- [21] FAN X, ZHANG F, SONG J, et al. A fine-grained policy model for provenance-based access control and policy algebras[J] 2020, arXiv: 2001.01945, 2023.
- [22] GROUP N B D P W, SUBGROUP D A T. NIST big data interoperability framework: volume 1, definitions[R]. 2019.
- [23] GROUP N B D P W, SUBGROUP D A T. NIST big data interoperability framework: volume 2, big data taxonomies[R]. 2019.
- [24] 李学龙, 龚海刚. 大数据系统综述[J]. 中国科学: 信息科学, 2015, 45(1): 1-44.
- LI X L, GONG H G. A survey on big data systems[J]. SCIENTIA SINICA Informationis, 2015, 45(1): 1-44.
- [25] ARSHAD H, JOHANSEN C, OWE O. Semantic attribute-based access control: a review on current status and future perspectives [J]. Journal of Systems Architecture, 2022, 129: 102625.
- [26] KAYES A S M, HAN J, COLMAN A. ICAF: a context-aware framework for access control[M]. Information Security and Privacy. 2012: 442-449.
- [27] KAYES A S M, RAHAYU W, WATTERS P, et al. Achieving security scalability and flexibility using fog-based context-aware access control[J]. Futur Gener Comp Syst, 2020, 107: 307-323.
- [28] KAYES A S M, HAN J, COLMAN A. An ontological framework for situation-aware access control of software services[J]. Information Systems, 2015, 53: 253-277.
- [29] MCINTOSH T, WATTERS P, KAYES A S M, et al. Enforcing situation-aware access control to build malware-resilient file systems[J]. Future Generation Computer Systems, 2021, 115: 568-582.
- [30] CORRADI A, MONTANARI R, TIBALDI D. Context-based access control for ubiquitous service provisioning[C]//Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04). 2004: 444-451.
- [31] BUI T, STOLLER S D, LI J J. Greedy and evolutionary algorithms for mining relationship-based access control policies[J]. Computers & Security, 2019, 80: 317-333.
- [32] IYER P, MASOUMZADEH A. Active learning of relationship-based access control policies[C]//Proceedings of the 25th ACM Symposium on Access Control Models and Technologies. 2020: 155-166.
- [33] KAYES A S M, HAN J, COLMAN A, et al. RelBOSS: a relationship-aware access control framework for software services [M]//MEERSMAN R, PANETTO H, DILLON T, et al. On the Move to Meaningful Internet Systems: Otm 2014 Conferences. 2014: 258-276.
- [34] SUN L, PARK J, NGUYEN D, et al. A provenance-aware access control framework with typed provenance[J]. IEEE Trans Dependable Secur Comput, 2016, 13(4): 411-423.
- [35] NGUYEN D, PARK J, SANDHU R. A provenance-based access control model for dynamic separation of duties[C]//Proceedings of 2013 Eleventh Annual Conference on Privacy, Security and Trust (PST). 2013: 247-256.
- [36] CHAKRABORTY S, SANDHU R. On feasibility of attribute-aware relationship-based access control policy mining[M]//Berlin: Springer. Data and Applications Security and Privacy. 2021: 393-405.
- [37] HU V C, FERRAILOLO D, KUHN R, et al. Guide to attribute based access control (abac) definition and considerations: NIST special publication 800-162[S]. 2014: 1-37.
- [38] KAYES A S M, HAN J, COLMAN A, et al. A semantic policy framework for context-aware access control applications[C]//Proceedings of 2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications. 2013: 753-762.
- [39] KAYES A S M, RAHAYU W, DILLON T. An ontology-based approach to dynamic contextual role for pervasive access control[C]//Proceedings 2018 IEEE 32nd International Conference on Advanced Information Networking and Applications. 2018: 601-608.
- [40] MOREAU L, CLIFFORD B, FREIRE J, et al. The open provenance model core specification (v1.1)[J]. Future Generation Computer Systems, 2011, 27(6): 743-756.
- [41] MISSIER P, BELHAJJAME K, CHENEY J. The W3C PROV family of specifications for modelling provenance metadata[C]//Proceedings of the 16th International Conference. 2013: 773-776.
- [42] BATRA G, ATLURI V, VAIDYA J, et al. Incremental maintenance of ABAC Policies[C]//Proceedings of the Eleventh ACM Conference on Data and Application Security and Privacy, 2021.
- [43] SHAN D, DU X, WANG W, et al. GNN-based method for predicting access control relationships for big data[C]//2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI). 2022.
- [44] HAN J, PEI J, YIN Y, et al. Mining frequent patterns without candidate generation: a frequent-pattern tree approach[J]. Data Mining and Knowledge Discovery, 2004, 8(1): 53-87.
- [45] HUANG H, FU Y, HU J, et al. Research on distributed dynamic trusted access control based on security subsystem[J]. IEEE Transactions on Information Forensics and Security, 2022: 1-15.
- [46] RUSHBY J. Noninterference, transitivity, and channel-control security policies[R]. 2005.
- [47] WANG X, JI H, SHI C, et al. Heterogeneous graph attention network[C]//Proceedings of The World Wide Web Conference(WWW '19). 2019: 2022-2032.
- [48] TU Z, LI R, LI Y, et al. Your apps give you away distinguishing mobile users by their app usage fingerprints[C]//Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies. 2018: 138:131-123.

[49] KARIMI L, ALDAIRI M. An automatic attribute based access control policy extraction from access logs[J]. IEEE Trans Dependable Secur Comput, 2022, 19(4): 2304-2317.

[50] SANDERS M W, YUE C, ACM. Mining least privilege attribute based access control policies[C]//Proceedings of 35th Annual Computer Security Applications Conference (ACSA). 2019: 404-416.



王文娟(1981-),女,河南鹤壁人,博士,信息工程大学教授,主要研究方向为云计算安全、入侵防御。

[作者简介]



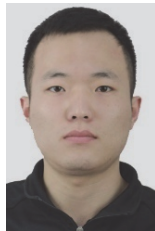
单捷斌(1982-),男,河北邯郸人,信息工程大学副教授,主要研究方向为大数据安全、信任安全、图神经网络。



王娜(1980-),女,山西临汾人,博士,信息工程大学副教授,主要研究方向为信息系统安全、大数据和区块链安全。



杜学绘(1968-),女,河南新乡人,博士,信息工程大学教授、博士生导师,主要研究方向为信息系统安全、大数据和区块链安全。



刘敖迪(1992-),男,黑龙江伊春人,博士,信息工程大学讲师,主要研究方向为大数据安全。