一种基于特征提取的访问控制方法

黄美蓉 欧 博 何思源
(湖南大学信息科学与工程学院 长沙 410082)

摘 要 当前,细粒度的授权控制是访问控制中的研究热点,它能够在单一固定的环境下合理地调整访问策略以满足工作流安全。然而,一旦其迁移到新场景,遭遇访问策略未设定的授权,它就可能难以给出正确判断,只能依靠人工审查来确认是否授权,但人工审查授权耗时耗力,在大数据环境下成本过高。因此,引入一种基于过去经验学习的自动化判别机制势在必行。文中尝试给出一种针对基于角色的多级访问控制模型的自动化审查方法,通过采样已有的正确和错误授权的时间、空间等特征来刻画出该访问控制的一般化特征表达,从而使得已有的访问控制模型在迁移环境下面对新情况依然能够给出正确判断,降低人工审查的工作量。实验表明,该分析机制对用户的访问请求有较高的正确评判率。

关键词 访问控制,多级授权管理,数据分析,特征
中图法分类号 TP309 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2019.02.017

Access Control Method Based on Feature Extraction

HUANG Mei-rong OU Bo HE Si-yuan
(College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China)

Abstract Recently, fine-grained authorization control has become a hot topic in access control research field, and it can adjust access strategy reasonably in a single fixed environment, so as to meet the safety of workflow. However, it may be difficult to give a correct judgement and only rely on manual checking to confirm whether it is authorized when it is migrated to the new scenario and encounters authorization that is not set by access policy. Manual checking is time-consuming, and it costs too much in big data environments. Therefore, it is imperative to introduce an automatic discrimination mechanism based on past experiences. This paper attempted to give an automatic discrimination method for role-based multilevel access control model, and described the general expression of the access control by sampling the correct and incorrect authorization time and space. This allows the existing access control model to make the righ judgements under the new environments, thus reducing the workload of manual review. The experimental results show that the analysis mechanism has a higher correct judge rate for user access requests.

Keywords Access control, Multi-level authorization management, Data analysis, Feature

1 引言

近年来,随着云计算、物联网等技术的快速发展,大数据时代也悄然来临。大数据分析技术在电子商务、医疗、金融等领域的广泛应用创造了巨大的社会价值;但大数据的安全隐私问题也阻碍着大数据的发展,例如"棱镜门"事件、Google发生的大规模用户资料外泄事件等,均造成了严重影响。因此,数据的安全和隐私问题成为现阶段研究的热点,访问控制技术是解决数据安全问题的重要手段之一。

最早出现的访问控制技术中,比较活跃的研究是自主访问控制(Discretionary Access Control, DAC)和强制访问控制(Mandatory Access Control, MAC)[1]。稍晚出现的基于角色的访问控制(Role-Based Access Control, RBAC)成为现在人们研究访问控制的奠基石。但这些研究都是基于被动型、静态性的访问控制,随着数据资源的计算与存储的多样化,人们从主动型、动态性、细粒度等方面着手研究,考虑任务、信任、时空、行为等角度[2-26],以满足不同的访问需求,使访问过程更加安全、灵活、高效。李昊等[2]总结出大数据访问控制技术的3个新特点:判定依据多元化、判定结果模糊化、多种访问控制技术融合化。Kuhlmann等[8]为了自动生成角色,将数据挖掘技术运用于RBAC中,并提出角色挖掘的概念,建立了一种"自下而上"的角色生成方法。Jafarian等[9]提出一种将角色挖掘问题转化为约束满足问题的通用方法。Martin等[10]使用数据挖掘的方法来检查访问策略中的一些错误漏洞,先自动生成请求,然后获取相应的响应,并在请求-响应上应用机器学习来推断策略属性。Jeffrey等[11]在基于行为的访问控制中运用机器学习的方法,分析企业人员的行为,并将分析结果作为访问控制的判断条件,首先对用户行为进行聚

到稿日期:2018-01-19 返修日期:2018-05-25 本文受国家自然科学基金-青年项目(61502160)资助。
黄美蓉(1991-),女,硕士生,主要研究方向为云环境下的访问控制;欧 博(1985-),男,博士,讲师,主要研究方向为信息隐藏、多媒体内容安全,E-mail:oubo@hnu.edu.cn(通信作者);何思源(1989-),男,硕士,主要研究方向为信息安全。

类,然后根据行为提取特征进行训练。马萌等[12]将基于条件随机场的机器学习用于优化 BLP 模型的规则,该方法对访问记录进行学习与训练,不但可根据当前系统的安全状态和安全事件动态调整安全规则,还可动态地限制敏感客体的读写范围。

在复杂多变的大数据背景下,工作流模式中设定的访问策略可能满足当前环境,但当这种访问策略迁移到新的访问环境且遭遇旧环境中未发生过的请求时,系统就可能无法对该新请求做出判断,甚至可能会做出错误判断,此时则需进行人工审查以确定该请求是否合乎规范,从而确定是否授权。当此类请求的数量巨大时,单靠人工审查将耗费大量的成本和时间。

为解决以上问题,文中提出在基于角色的多级访问控制模型上增加一个数据分析机制,通过这样一个机制可以预测访问请求的合理性。如访问请求合理,则将其交由系统的多级访问控制策略来进行进一步判断;如访问请求不合理,则直接拒绝该访问。通过决策分析以及机器学习分析两种方法来分析用户的访问请求,从而提高系统的授权效率和安全性。

本文第 2 节提出基于角色的多级访问控制模型,定义了基本元素的概念,并描述了多级安全策略以及访问控制规则;第 3 节介绍了基于规则的行为特征提取;第 4 节进行实验并分析实验结果。最后总结全文。

2 基于角色的多级访问控制模型

本文提出的基于角色的多级访问控制模型结合了 RBAC 和 MAC 两种访问控制模型,既具有 RBAC 的灵活性,又体现了 MAC 的高安全性。该模型能解决工作流模式在时间、空间、安全级别等方面都有较高要求的细粒度多级授权管理问题。为了进一步提高工作流模式中的授权效率,本文在多级安全模型上增加了一种数据分析机制,以分析历史访问的数据记录,从而预测评判访问请求是否合乎规范。基于角色的多级访问控制模型如图 1 所示。

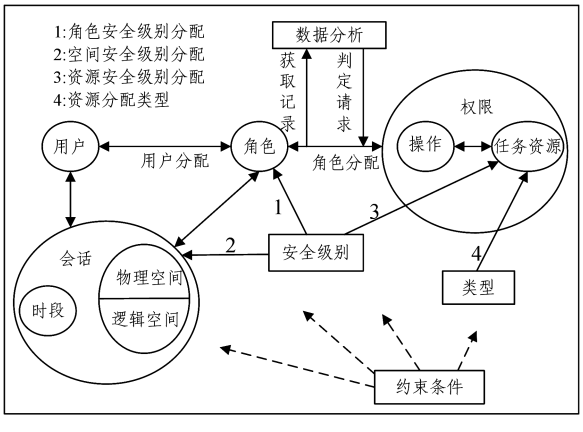

图 1 基于角色的多级访问控制模型

Fig. 1 Role-based multilevel access control model

2.1 基本元素的概念

角色:实现某些功能所需的权限集合。本文将角色划分成不同安全访问等级,因此角色表示成一个二元组,即角色=(角色,角色安全级别),$R=\{(r_i,\alpha(r_i))\mid i\in N\}$,其中 r_i 表示角色,$\alpha(r_i)$表示角色安全级别,并且 α 之间存在一种偏序关系≥。给角色增加安全等级标识,其与 MAC 中的主体相一致。基于此,融合 RBAC 与 BLP,实现多级授权管理。

会话:用户对系统发送访问请求时建立会话,通过会话激活相关角色集中的一个子角色集,然后将其指派给用户。用户通过会话获得的权限就是该会话激活的所有角色所拥有的权限的并集。会话期间,会判定角色所处的时段和空间。

权限:角色对资源的操作集合,$P=\{p_i\mid i\in N\}$。权限分为读权限、写权限和执行权限,以及通过这些权限丰富起来的一系列权限。

任务资源:角色操作的对象集合。本文将任务资源定义为一个三元组,即任务资源=(任务,资源安全级别,资源类型),$S=\{(s_i,\beta(s_i),\lambda(s_i))\mid i\in N\}$。其中 s_i 表示任务;$\beta(s_i)$表示资源安全级别,并且 β 之间存在一种偏序关系≥;$\lambda(s_i)$表示资源类型。

时段:表示时间段的约束集合,$T=\{t_i\mid i\in N\}$。为便于区分,用 $T(R)$表示角色允许使用的时段,用 $T(S)$表示任务资源允许访问的时段。

逻辑空间:表示网络逻辑位置所构成的集合。我们将逻辑空间定义成一个二元组,即逻辑空间=(网络逻辑位置,空间安全级别)。空间层次越机密,安全级别就越高。$Lp=\{(lp_i,\gamma_j)\mid i,j\in N\}$,其中,$lp_i$ 表示网络的逻辑位置,γ_j 表示空间安全级别,并且 γ 之间存在一种偏序关系≥。用 $Lp(R)=\{(lp(r_i),\gamma(r_i))\mid i\in N\}$表示角色 R 所处的逻辑空间,用 $Lp(S)=\{(lp(s_i),\gamma(s_i))\mid i\in N\}$表示任务资源 S 所处的逻辑空间。

物理空间:表示网络物理位置所构成的集合。$Lo=\{lo_i\mid i\in N\}$,其中 lo_i 表示某一处场所的网络物理位置。用 $Lo(R)=\{lo(r_i)\mid i\in N\}$表示角色 R 所处的物理空间,用 $Lo(S)=\{lo(s_i)\mid i\in N\}$表示任务资源 S 所处的物理空间。

这里的逻辑空间主要是指主机的 MAC 地址和 IP 地址。逻辑空间与物理空间之间相互联系,都属于空间的范畴,本文只对逻辑空间划分安全级别,对物理空间则不再赘述。

2.2 多级安全策略描述

在多级安全访问控制中,需要满足多级安全访问策略。本文结合角色安全级别、物理空间(范畴)、逻辑空间安全级别、任务资源安全级别和类型(范畴)以及时段限制对多级安全访问策略进行相关描述。

定义安全标记函数 $F=L\times C\times T$,其中 F 表示安全标记,L 表示安全级别,C 表示范畴,T 表示时段。本文中 $L=(\alpha(r),\beta(s),\gamma(r),\gamma(s))$,$C=(lo(r),lo(s),\lambda(s))$,$T=(t(R),t(S))$。安全标记函数体现的是一种多级安全策略。假设 r 对 s 请求访问,如果 r 和 s 之间的安全级别、范畴和时段的关系用 f 表示,其符合安全标记函数 F 的描述,用数学符号表示为 $f\in F$,则允许 r 对 s 的访问,否则拒绝。设有安全标记 $F1=L1\times C1\times T1$,$F2=L2\times C2\times T2$。当且仅当 $L1\geqslant L2$,$C1\supseteq C2$,$T1\supseteq T2$ 同时成立,有 $F1\geqslant F2$,表示 $F1$ 支配 $F2$;当且仅当 $L1>L2$,$C1\supset C2$,$T1\supset T2$ 同时成立,有 $F1>F2$,表示 $F1$ 真支配 $F2$;当且仅当 $L1=L2$,$C1=C2$,$T1=T2$ 同时成立,有 $F1=F2$,表示 $F1$ 等于 $F2$。

约束 1 在低安全级别的空间,不允许读和执行高安全级别的任务资源,即 $L(\gamma(S))\geqslant L(\beta(S))$;在高安全级别的空

间不允许写低安全级别的任务资源，即 $L(\beta(S))\geqslant L(\gamma(S))$。图 2 显示了逻辑空间与任务资源之间的约束关系。

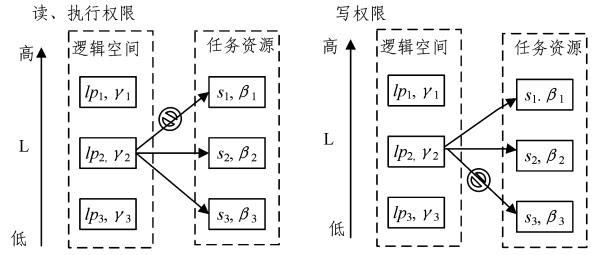

图 2　逻辑空间与任务资源之间的读、执行和写约束

Fig. 2　Reading, execution and writing constraints between logical space and task resources

约束 2　在高安全级别的空间不允许使用低安全级别的角色，即 $L(\alpha(R))\geqslant L(\gamma(R))$。图 3 显示了角色与逻辑空间之间的约束关系。

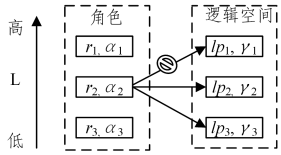

图 3　角色与逻辑空间之间的约束

Fig. 3　Constraints between role and logical space

约束 3　高安全级别的角色在低安全级别的空间可以读和执行低安全级别的任务资源，可以写高安全级别的任务资源。图 4(a)和图 4(b)显示了角色、逻辑空间、任务资源三者之间的读、执行和写的约束。

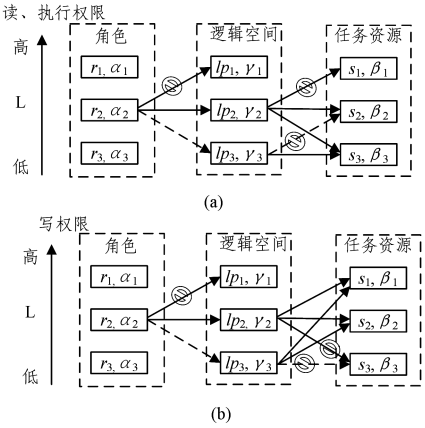

图 4　角色、逻辑空间、任务资源三者之间的读、执行和写约束

Fig. 4　Reading, execution and writing constraints among roles, logical space and task resources

如图 4(a)中的虚线路径所示，如果任务资源的安全级别不高于角色的安全级别，但不低于空间的安全级别，那么该角色不能在该空间读和执行该任务资源。如图 4(b)中的虚线路径所示，如果任务资源的安全级别不低于空间的安全级别，但不高于角色的安全级别，那么该角色不能在该空间写该任务资源。这个约束说明如果高安全级别的角色读和执行低安全级别的任务资源，那么需要优先满足约束 1；如果低安全级别的角色写高安全级别的任务资源，那么需要优先满足 BLP 模型的“上写”特性。

约束 4　角色访问不同类型的任务资源必须在相应的指定空间进行，即 $C(Lo(S))\subseteq C(Lo(R))$。如图 5 所示，角色想要访问高安全性类型的资源，就必须去相应高安全性（安全级别高）的空间。

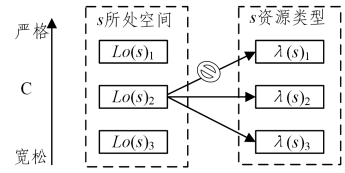

图 5　物理空间与任务资源类型之间的约束

Fig. 5　Constraints between physical space and task resource types

约束 5　权限只能在角色和任务资源许可的时段内使用，即 $T(t(S))\supseteq T(t(R))$。角色在允许激活的时段内，可以激活（不一定能使用）该权限，不同安全级别的角色，其激活时段不同。任务资源在允许访问的时段内，角色才可以使用该权限，不同安全级别、不同类型的任务资源，其访问时段不同。只有任务资源允许访问的时段包含了角色允许激活的时段，才可以使用相关权限。

当用户向系统发送访问请求，请求操作任务资源时，系统建立一个会话，检查用户在该会话中所激活的子角色集的权限是否能够操作该任务资源，检查角色的安全级别、任务资源的安全级别、逻辑空间的安全级别和任务资源的物理空间是否包含角色所处的物理空间，以及这些空间之间的关系是否满足安全标记函数的要求，检查用户激活的角色和任务资源是否在允许的时段内进行操作，如果上述条件均满足，则允许用户访问，否则系统将拒绝用户访问请求。

2.3　相关访问规则

本文根据用户对数据资源的操作，结合 BLP 模型的系统安全规则。现给出如下访问控制规则。

规则 1　自主安全规则

参考 BLP 模型中，系统状态 v 由一个有序三元组 $(b\times M\times F)$ 表示，v 满足自主安全规则，当且仅当 $(b=(r,s,p))\in(B=(R,S,P))$，$P\in M$。

其中，b 表示在 v 状态下，某些级别的角色 r 通过某些访问权限 p 来操作某些任务资源 s，v 状态体现了时段、逻辑空间和物理空间元素。M 表示访问控制矩阵，包含了所有角色对所有任务资源的访问权限，M 中的元素 $m_{ij}\in P$，表示角色 r_i 操作任务资源 s_j 所对应的权限，m_{ij} 也有可能表示为空。F 表示安全标记函数，2.2 节已对 F 进行了详细阐述。

规则 2　读（执行）权限规则

角色 r 在系统状态 $v=(b\times M\times F)$ 下对任务资源 s 申请读（执行）权限，则 r 可读（执行）安全级别不大于 $L(\gamma(s))$，且任务资源类型不超过当前 $\lambda(s)$ 的最大类型的 s，当且仅当 $T(t(s))\supseteq T(t(r))$，且 $L(\alpha(r))\geqslant L(\gamma(r))=L(\gamma(s))\geqslant L(\beta(s))$，$C(lo(r))=C(lo(s))\supseteq C(\lambda(s))$。

该规则说明角色处于不高于其安全级别的空间时，可以在该空间内对不大于该空间安全级别的资源进行读（执行）操作，任务资源的安全级别不大于该空间安全级别，该条件满足 2.2 节中描述的约束。对于其他情况，r 对 s 不可读（执行）。

规则 3　写权限规则

角色 r 在系统状态 $v=(b\times M\times F)$ 下对任务资源 s 申请

写权限，则 r 的可写安全级别不小于 $L(\gamma(s))$，且任务资源类型不低于当前 $\lambda(s)$ 的最大类型的 s，当且仅当 $T(t(s))\supseteq T(t(r))$，且 $L(\beta(s))\geqslant L(\alpha(r))\geqslant L(\gamma(r))=L(\gamma(s))$，$C(lo(r))=C(lo(s))\supseteq C(\lambda(s))$。

该规则说明角色处于不高于其安全级别的空间时，可以在该空间内对不小于该空间安全级别的资源进行写操作，该条件满足 2.2 节中描述的约束。对于其他情况，r 对 s 不可写。

现对读权限规则和写权限规则进行举例说明，执行权限规则与读权限规则相同。如图 6 所示，假设安全模型中角色、空间和任务资源分别有 3 个安全级别(1,2,3)，任务资源的类型有 3 类(1,2,3)。图 6(a)中角色 r 在 $L(\alpha(r))=1$、空间 $L(\gamma(r))=L(\gamma(s))=1$、任务资源 $L(\beta(s))=1$ 时可读的资源数有 3 种；以此类推，角色 r 在 $L(\alpha(r))=1$、空间 $L(\gamma(r))=L(\gamma(s))=1$ 时可读的资源数有 6 种；角色在 $L(\alpha(r))=1$ 时可读的资源数有 10 种。若 r 与 s_1 存在如下关系：$T(t(s_1))\supseteq T(t(r))$，$C(lo(r))=C(lo(s_1))\supseteq C(\lambda(s_1))$，$L(\alpha(r))>L(\gamma(r))=L(\gamma(s_1))=L(\beta(s_1))$，则 r 对 s_1 进行读访问，如图 6(a)虚线所示。图 6(b)中角色 r 在 $L(\alpha(r))=2$、空间 $L(\gamma(r))=L(\gamma(s))=2$、任务资源 $L(\beta(s))=1$ 时可写的资源数有 3 种；以此类推，角色 r 在 $L(\alpha(r))=2$、空间 $L(\gamma(r))=L(\gamma(s))=2$ 时可写的资源数有 5 种；角色在 $L(\alpha(r))=2$ 时可写的资源数有 10 种。若 r 与 s_2 存在如下关系：$T(t(s_2))\supseteq T(t(r))$，$C(lo(r))=C(lo(s_2))\supseteq C(\lambda(s_2))$，$L(\beta(s_2))>L(\alpha(r))>L(\gamma(r))=L(\gamma(s_2))$，则 r 对 s_2 进行写访问，如图 6(b)虚线所示。

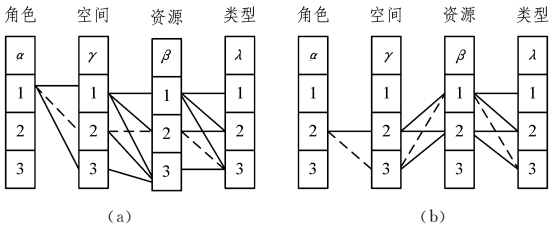

图 6 读权限与写权限规则示例

Fig. 6 Example of read permission and write permission rules

3 基于规则的行为特征提取

在用户访问资源的过程中，会产生大量的访问数据及响应数据。这些数据中大部分都是合乎规则的，体现了访问控制规则。根据访问控制所必须的要素可以提炼出一些访问特征，文中列出一些值得关注的特征，当然也不限于本文所列，根据不同的访问规则或控制模型，提取的特征也会有所不同。

访问控制中主体可以考虑作为一个特征，这是一个重要要素，有了主体，才能对客体进行访问。主体又可以细分为很多小特征，除了考虑主体本身外，还有主体的安全级别也可以作为特征。主体拥有的权限也可以作为特征。

考虑主体作为特征后，就要考虑客体作为特征。客体是主体访问的对象，可以是各种资源、任务等。除了客体自身外，客体的类型也是一类特征。客体的安全级别也需要考虑。

但上述因素都是静态性的，访问控制的发展需要考虑更多灵活的因素，如时间、空间上的相关考虑。从时间上来说，主体对客体拥有的权限并不是一成不变的。时间是一个很重要的特征，可以考虑将访问发生的开始时间和结束时间作为特征，或者是发生同一种访问的某一个时间段、上班时间、下班时间、工作日、休息日等作为特征，然后基于这些信息提取数据。也可以从访问发生的空间角度来考虑特征，对于一些访问要求较高的单位，如军工或者是保密性较高的政府部门、企业等都可以考虑空间要素。空间里又分很多小特征，比如物理空间和逻辑空间，以及将计算机的软硬件平台作为特征。空间的范畴也可以考虑，在比较大的空间范畴可以访问安全性要求不高的客体。空间的安全级别也是一类特征，对于不同安全级别的空间，访问的主体的安全级别要与之对应，访问的客体的安全级别也要与之对应。

本文根据提出的模型提取一些特征，对于主体选择角色、角色安全级别、角色行使的权限 3 个特征，对于客体选择任务资源类型和任务资源的安全级别 2 个特征，时间特征选择时段，空间特征选择物理空间的范畴、逻辑空间的安全级别 2 个特征，共选择了 8 个特征。根据这 8 个特征提取出一些数据，然后对数据进行实验并分析。

在进行实验建模时，通过采集系统的历史记录并分析数据特征，将用户申请角色发送的访问请求作为输入，将系统对该请求作出的决策响应作为输出。在用户的访问记录中，收集到的特征向量是上述的 8 个特征，其取值及意义如表 1 所列。

表 1 特征取值及意义

Table 1 Feature value and significance

特征	取值	意义
角色	1～16	根据系统设置的角色数量来确定
角色行使的权限	1～48	其中 1～16 是对应角色的读权限，17～32 是角色的执行权限，33～48 是角色的写权限
时段	1～6	将一天 24 h 划分为 6 个时段，其中 22 点到 8 点是时段 1，8 点到 12 点是时段 2，12 点到 14 点是时段 3，14 点到 18 点是时段 4，18 点到 19 点是时段 5，19 点到 22 点是时段 6
物理空间	1～11	根据系统设置的地理位置数量来确定，其中 11 表示不确定的地理位置
空间安全级别	1～6	根据系统设置的空间安全级别来确定，其中 6 表示不确定的空间安全级别
资源安全级别	1～4	根据系统设置的资源安全级别确定
角色安全级别	1～5	根据系统设置的角色安全级别确定
资源类型	1～10	根据系统设置的资源类型确定

将上述的特征向量作为输入，将系统的响应作为输出。系统允许用户请求则输出响应 1，表示符合访问规则；拒绝则输出响应－1，表示不符合访问规则。对每条历史记录进行预处理后都会得到一个 8 维的特征向量，如某条记录的对应向

量为(7,23,2,4,3,3,4,4)，此条数据的第 1 列表示角色序号为 7，第 2 列表示该角色申请与其对应的执行权限为 23，第 3 列表示用户在 8 点到 12 点申请该请求，第 4 列表示角色处于地理位置为 4 的地方，第 5 列表示该地理位置的空间安全级别为 3，第 6 列表示用户申请执行资源任务的安全级别为 3，第 7 列表示该角色的安全级别为 4，第 8 列表示用户访问的资源类别为 4。在训练过程中将系统的响应放在数据之前，这条数据的响应为 1，表示系统允许该请求。通过如此处理，我们可以清楚地知道用户访问请求的相关信息，从而预测判断系统允许或拒绝该访问请求。

4　实验分析

根据本文基于角色的访问控制模型模拟某公司的访问情况，通过历史记录提取数据，把获得的 12500 条数据标注为 1 和－1 两类，1 表示正确的数据，也即该数据符合访问策略；－1 表示错误的数据，也即不符合访问规则的访问记录。表 2 列出了本文实验数据的分布比例情况，其中标注为 1 的数据有 10372 条，标注为－1 的数据有2128 条，读权限的数据有 5943 条，执行权限的数据有 4343 条，写权限的数据有 2214 条。采用 libSVM 进行实验测试，并采用两种交叉验证法来评估实验结果。这两种交叉验证方法分别为 2 折交叉和 K 折交叉。

表 2　数据的分布比例

Table 2　Distribution proportion of data

标签	权限			总数
	读	执行	写	
1	4724	3870	1778	10372
－1	1219	473	436	2128
总数	5943	4343	2214	12500

本文中 2 折交叉是将数据集随机分成两份，一份用作训练集，另一份用作测试集，训练集占总数据集的 70%，测试集占总数据集的 30%。实验重复 3 次后取平均值。总数据集有以下 4 种取法：读和执行权限的数据(数据集 1)，读和写权限的数据(数据集 2)，执行和写权限的数据(数据集 3)，读、执行和写权限的数据(数据集 4)。采用这 4 种取法分别进行 4 次测试，并取平均值。

K 折交叉是将数据集随机分成 K 份，其中第 K 份作为测试集，剩余(K－1)份作为训练集，共迭代 K 次，此处 K 值为 5。本文将训练集进行 5 折交叉验证，其中 4 份用作训练集，1 份用作验证集，通过 5 次迭代，选择最好的一次迭代结果，然后调用模型参数建模。最后将测试集作为输入在此模型上进行测试，得出预测结果。表 3 列出了 4 种总数据集的训练集、验证集、测试集的分配情况。

表 3　不同数据集的数据分配情况

Table 3　Data distribution of different datasets

数据集	权限			总数
	训练集	验证集	测试集	
数据集 1	7200/5760	7200/1440	3086	10286
数据集 2	5710/4568	5710/1142	2447	8157
数据集 3	4590/3672	4590/918	1967	6557
数据集 4	8750/7000	8750/1750	3750	12500

表 4 列出了交叉验证的综合结果，由表中结果可以看出，在访问控制中进行特征提取，对访问历史记录进行学习训练，可对访问数据进行较为准确的预测，平均能获得接近 98%的正确率。从表 4 中还可以看出，对于标签 1，也即合乎规范的访问记录来说，预测得到的平均准确率接近 98%，平均召回率接近 100%，取得了较好结果。$F1$ 值为一个综合评价指标，综合了准确率和召回率的结果。标签为 1 的 $F1$ 值达到将近 99%，这说明本文实验方法非常有效。

表 4　交叉验证结果

Table 4　Cross-validation results

(单位：%)

总数据集	标签	准确率	召回率	$F1$ 值	平均值
数据集 1	1	97.13	99.60	98.35	97.11
	－1	99.70	85.04	91.79	
数据集 2	1	98.25	99.41	98.83	99.91
	－1	99.05	89.00	93.76	
数据集 3	1	97.95	99.15	98.55	97.11
	－1	96.24	85.45	90.52	
数据集 4	1	97.51	99.66	98.57	97.52
	－1	98.50	84.54	90.99	
平均值	1	97.71	99.46	98.57	97.92
	－1	98.37	86.01	91.76	

结束语　本文首先描述了一个基于角色的多级访问控制模型，实现了 RBAC 模型与 BLP 模型的有机结合，解决了用户不同时间、不同空间、不同安全级别的多级授权管理问题，从而提高了系统的安全性。在工作流模式中，由于系统难以事先设定满足所有情况的细粒度访问策略，出现此类情况需要进行人工审查，这将耗费大量的成本和时间。因此，我们在该模型上增加了一个数据分析机制以分析历史访问数据，并根据该机制评判访问请求是否合乎规范。对合乎规范和不合乎规范的访问数据进行提取分析，然后通过交叉验证技术建立一个较好的实验模型。实验表明，该数据分析机制对用户的访问请求有较高的正确评判率，在访问控制中结合机器学习的方法有较好的参考价值和实际应用前景。

参 考 文 献

[1]　WANG Y D，YANG J H，XU C，et al. Survey on access control technologies for cloud computing[J]. Journal of Software，2015，26(5)：1129-1150. (in Chinese)
王于丁，杨家海，徐聪，等. 云计算访问控制技术研究综述[J]. 软件学报，2015，26(5)：1129-1150.

[2]　LI H，ZHANG M，FENG D G，et al. Research on access control of big data[J]. Chinese Journal of Computers，2017，40(1)：72-91. (in Chinese)
李昊，张敏，冯登国，等. 大数据访问控制研究[J]. 计算机学报，2017，40(1)：72-91.

[3]　UZUN E，ATLURI V，SURAL S，et al. Analyzing temporal role based access control models[C]// Proceedings of the 17th ACM symposium on Access Control Models and Technologies. ACM，2012：177-186.

[4]　RANISE S，TRUONG A，ARMANDO A. Scalable and precise automated analysis of administrative temporal role-based access control[C]// Proceedings of the 19th ACM Symposium on Ac-

cess Control Models and Technologies. ACM,2014:103-114.

[5] BERTINO E,CATANIA B,DAMIANI M L,et al. GEO-RBAC:A spatially aware RBAC[C]// Proceedings of the 10th ACM Symposium on Access Control Models and Technologies. New York:ACM Press, 2005:29-37.

[6] ANDROULAKI E,SORIENTE C,MALISA L,et al. Enforcing location and time-based access control on cloud-stored data[C]// 2014 IEEE 34th International Conference on Distributed Computing Systems (ICDCS). IEEE,2014:637-648.

[7] LI F H, WANG W, MA J F,et al. Action-based access control model and administration of actions[J]. Acta Electronica Sinica, 2008,36(10):1881-1890. (in Chinese)
李凤华,王巍,马建峰,等. 基于行为的访问控制模型及其行为管理[J]. 电子学报,2008,36(10):1881-1890.

[8] KUHLMANN M,SHOHAT D,SCHIMPF G. Role mining-revealing business roles for security administration using data mining technology[C]// Proceedings of the Eighth ACM Symposium on Access Control Models and Technologies. ACM, 2003:179-186.

[9] JAFARIAN J H,TAKABI H,TOUATI H,et al. Towards a general framework for optimal role mining:A constraint satisfaction approach[C]// Proceedings of the 20th ACM Symposium on Access Control Models and Technologies. ACM,2015:211-220.

[10] MARTIN E,XIE T. Inferring access-control policy properties via machine learning[C]// Seventh IEEE International Workshop on Policies for Distributed Systems and Networks. IEEE, 2006.

[11] CLEVELAND J, MAYHEW M J, ADLER A,et al. Scalable machine learning framework for behavior-based access control [C]// 2013 6th International Symposium on Resilient Control Systems (ISRCS). IEEE,2013:181-185.

[12] MA M,TANG Z,LI R F,et al. Improved BLP Model Based on CRFs[J]. Computer Science,2015,42(8):138-144,151. (in Chinese)
马萌,唐卓,李仁发,等. 基于条件随机场的改进型 BLP 访问控制模型[J]. 计算机科学,2015,42(8):138-144,151.

[13] CRAMPTON J,MORISSET C,ZANNONE N. On missing attributes in access control:Non-deterministic and probabilistic attribute retrieval[C]// Proceedings of the 20th ACM Symposium on Access Control Models and Technologies. ACM, 2015:99-109.

[14] LI J,SQUICCIARINI A,LIN D,et al. Secloc:securing location-sensitive storage in the cloud[C]// Proceedings of the 20th ACM Symposium on Access Control Models and Technologies. ACM,2015:51-61.

[15] JAYARAMAN K,GANESH V,TRIPUNITARA M,et al. Automatic error finding in access-control policies[C]// Proceedings of the 18th ACM Conference on Computer and Communications Security. ACM,2011:163-174.

[16] OH S,PARK S. Task-role-based access control model[J]. Information Systems,2003,28(6):533-562.

[17] ARDAGNA C A,CREMONINI M,DAMIANI E,et al. Supporting location-based conditions in access control policies[C]// Proceedings of the 2006 ACM Symposium on Information,Computer and Communications Security. ACM,2006:212-222.

[18] RAY I,KUMAR M,YU L. LRBAC:a location-aware role-based access control model[C]// International Conference on Information Systems Security. Springer Berlin Heidelberg, 2006:147-161.

[19] RAY I,TOAHCHOODEEM. A spatio-temporal role-based access control model[C]// IFIP Annual Conference on Data and Applications Security and Privacy. Springer Berlin Heidelberg, 2007:211-226.

[20] CHEN H C,WANG S J,WEN J H,et al. Temporal and Location-based RBAC model[C]// Fifth International Joint Conference on INC,IMS and IDC. IEEE,2009:2111-2116.

[21] CHAKRABORTY S,RAY I. TrustBAC:integrating trust relationships into the RBAC model for access control in open systems[C]// Proceedings of the Eleventh ACM Symposium on Access Control Models and Technologies. ACM,2006:49-58.

[22] LANG B. Access control oriented quantified trust degree representation model for distributed systems[J]. Journal on Communications,2010,31(12):45-54. (in Chinese)
郎波. 面向分布式系统访问控制的信任度量化模型[J]. 通信学报,2010,31(12):45-54.

[23] KANDALA S,SANDHUR. Secure role-based workflow models [M]// Database and Application Security XV. Springer US, 2002:45-58.

[24] BOTHA R A,ELOFF J H P. Designing role hierarchies for access control in workflow systems[C]// Computer Software and Applications Conference,2001(COMPSAC 2001). IEEE,2001: 117-122.

[25] SUN Y,MENG X,LIU S,et al. Flexible workflow incorporated with RBAC[C]// International Conference on Computer Supported Cooperative Work in Design. Springer Berlin Heidelberg, 2005:525-534.

[26] YAO H B,HU H P,LU Z D,et al. Dynamic role and context-based access control for grid applications[J]. Computer Science, 2006,33(1):41-44. (in Chinese)
姚寒冰,胡和平,卢正鼎,等. 基于角色和上下文的动态网格访问控制研究[J]. 计算机科学,2006,33(1):41-44.