基于机器学习的企业私有云用户行为分析模型

郑明辉,吕经华
(湖北民族学院 信息工程学院,恩施 445000)

摘 要 鉴于传统的访问控制机制为用户提供静态的授权,极少监管用户行为,导致企业很难发现来自恶意的内部用户或账号失窃带来的数据安全威胁,定义和形式化描述了私有云下的用户行为,基于 Hadoop 架构提出了一个基于 TensorFlow 机器学习的用户行为分析的框架,通过用户行为数据采集、存储、特征选择、预处理,给出了建立并训练用于用户行为分析的机器学习模型的方法和过程,实现了企业私有云用户行为的自动分析,用于数据安全威胁的发现和响应.

关键词 私有云;数据安全;用户行为;机器学习
中图分类号 TP393.2 **文献标识码** A **文章编号** 1672-4321(2017)03-0095-06

A User Behavior Analysis Model Based on Machine Learning for Enterprise Private Cloud

Zheng Minghui, *Lü Jinghua*
(College of Information Engineering, Hubei University for Nationalities, Enshi 445000, China)

Abstract Since traditional access control mechanisms provide static authorization for users, but with little supervision over their behaviors, it is difficult for enterprises to locate threats of data security posed by malicious insiders or compromised user accounts with high privilege. This article defines and formalizes user behaviors under Private Cloud, Through user behavior data acquisition, storage, feature selection and preprocessing, A Neural Network model for User Behavior Analysis is proposed with its training and optimization procedure. Thus a User Behavior Analysis model based on TensorFlow Machine Learning platform over Hadoop framework is given, which can analysis user behaviors automatically, and can help enterprises to locate and respond to threats of data security efficiently in private cloud.

Keywords Private Cloud; Data Security; User Behavior; Machine Learning

近年来,从大量报道的信息安全事件中得知,很多造成重大影响的数据安全事件,都是由于企业内部账号,尤其是高权限账号的泄漏所导致的[1]. 可见,企业把信息系统集成整合到私有云后,多样化的攻击方式使企业数据资源面临着更加复杂的威胁.

云计算广泛应用后,信息安全进入了立体和深度防御的时代,形成了以预警、攻击防护、响应恢复为主的全生命周期安全管理[2]. 基于属性的访问控制模型(ABAC)较为适用于云环境下细粒度、多层次的访问控制需求[3,4],但基于角色的访问控制模型也会在企业私有云中继续存在一段时间. 传统的访问控制机制通常不会对已获授权的合法用户行为进行动态监管,因此企业对数据泄漏和攻击行为的发现普遍滞后[5]. 考虑到任何一种攻击都是由一系列的信息系统用户行为实现的[6],人们引入用户行为分析(UBA)从大量用户行为中,找出潜在的攻击行为,及时预警,指导企业的安全团队及时响应潜在威胁[7]. 把用户行为分析与访问控制模型结合起来,可以实现基于信任的动态授权和访问控制,有助于保障数据安全和隐私[8-10],而机器学习的方法近年来被广泛应用于用户行为分析,可以推动云计算环境下的数据安全保障[11,12]. 但很多宣称基于机器学习的安全产品仍然是基于静态规则进行判断的,同样无法有效识别新的恶意行为,以此为基础的入侵检测、入侵防御、防火墙和信息安全软件,每天都会产生数以万计的信息安全事件信息,混杂着大量误报和漏报,已经超越了人工

收稿日期 2017-05-17
作者简介 郑明辉(1972-),男,教授,博士,主要研究方向为信息安全,E-mail: mhzheng3@163.com
基金项目 国家自然科学基金资助项目(61472121)、湖北省创新群体项目(2016CFA021)

分析的极限,使安全从业者无所适从. 传统的机器学习算法通常基于有限的训练数据进行学习,不适合在云环境下分析持续产生的、具有大数据特征的用户行为.

随着近年来机器学习技术的快速发展,国际上使用机器学习进行用户行为分析的技术在数据安全领域的迅速推广[13-16],催生了一些基于机器学习的用户行为分析安全产品,包括 Exabeam、Fortscale、Niara、Secronix、Splunk 等. 这些产品使用机器学习算法动态分析用户行为,侦测用户行为变化,发现潜在的威胁,可以定位恶意的内部员工和外部攻击者,能实现大量安全事件的自动分析,为企业的信息安全团队提供有用的建议. 但这些产品大多用于单机或单独的信息系统,不能直接应用于企业私有云下多个信息系统集成整合环境下的用户行为分析.

本文在企业私有云环境下,首先对用户行为进行定义和形式化,设计了一个基于机器学习的用户行为分析模型,随后基于 Hadoop 架构,提出了使用 TensorFlow 机器学习开源平台进行用户行为分析的方法和实现过程,自动分析企业私有云环境下大量信息系统用户行为数据,找出异常的用户行为,保障企业的数据安全.

1 企业私有云用户行为定义和形式化描述

首先给出企业私有云环境下的用户行为定义,并进行形式化描述.

(1) 企业私有云下的用户行为环境. 假设企业私有云的基础设施资源池中有 l 台物理服务器,记为 PS $[ps,ps_1,ps_2,\cdots,ps_s,\cdots,ps_l]$,按需建立了 m 个虚拟机,记为 VM $[vm_1,vm_2,\cdots,vm_i,\cdots,vm_m]$,每个虚拟机安装了相应的操作系统,记为 OS $[os_1,os_2,\cdots,os_i,\cdots,OS_m]$,运行着企业的 i 个数据库,记为 DB $[db_1,db_2,\cdots,db_i]$,j 个中间件,记为 MW $[mw_1,mw_2,\cdots,mw_j]$,以及 k 个信息系统,记为 S $[s_1,s_2,\cdots,s_k]$ 形成企业信息系统集成整合平台. 在企业的员工集合的某个子集中,每个员工拥有唯一的账号,通过单点登录后,可以访问获得授权的多个信息系统,用户账号的集合记为 U $[u_1,u_2,\cdots,u_n]$.

(2) 用户行为的定义和形式化描述. 用户行为(User Behavior,UB):是指某个用户访问企业数据资源的全过程. 一个完整的用户行为可由行为时态(Behavior Environment,BE)、(Behavior Subject,BS)、行为客体(Behavior Object,BO)、行为操作(Behavior Operation,BP)等属性集来描述,用户行为的集合可以记为: UB $[BE,BS,BO,BP]$.

当某个员工使用某个账号 u_i 登录某个信息系统 s_i 后,根据岗位、职责、信任等级、时间、地点、终端类型等属性子集,访问控制模块为其在角色集合 R $[r_1,r_2,\cdots,r_i,\cdots,r_n]$ 中选择适合的角色,从一组操作权限集合 P $[P_1,P_2,\cdots,p_i,\cdots,p_n]$ 中获得一个权限子集 P',获准访问信息系统模块 M $[m_1,m_2,\cdots,m_i,\cdots,m_n]$ 和功能集合 F $[f_1,f_2,\cdots,f_i,\cdots,f_n]$ 的某个子集,从而访问数据库表 TB $[tb_1,tb_2,\cdots,tb_i,\cdots tb_n]$ 的某些子集,或者按某些属性 Att $[a_1,a_2,\cdots,a_i,\cdots,a_n]$ 访问数据记录,这个过程被称为一次用户行为. 记为 ub_i $[be_i,bs_i,bo_i,bp_i]$.

下面分别定义用户行为的时态、主体、客体和操作,从而形式化描述每一次用户行为的详细特征.

行为时态:是指用户行为发生的时态信息,包括行为的时间、地点、终端类型等信息. 一次用户行为的时态记为 be_i $[Ts_i,Te_i,Loc_i,Terminal_i]$,其中 Ts_i,Te_i 代表行为开始和结束的时间,取自日志中记录的时间. 代表用户发起行为的地点,以 IP 地址或终端所在的地址描述,表示用户发起行为的终端类型,取自日志记录.

行为主体:是指任何发起访问数据的实体. 一个员工账号作为行为主体时,可以表示为:

bs_i $[UserName,\ AccountID,\ UserCreatedTime,\ System,\ UserClassfication,\ UserRisk,\ UserTrustLevel,\ UserRole,\ BehaviorCount]$ 这些属性描述了用户的姓名、账号、创建时间、所属系统、分类、风险分数、信任等级、角色、行为次数等 $Terminal_i$,用来刻画一个行为主体的特征.

行为客体:是指行为主体可以访问的数据对象. 把一次用户行为访问到的企业数据集及其属性表示为一个九元组,记为 bo_i $[ps_i,vm_i,os_i,s_i,m_i,f_i,db_i,tb_i,a_i]$,表示每一次访问的行为客体集. 统计每个实体 e_i 被访问的次数 x_{e_i},可计算每个实体被用户访问到的概率 p_{e_i},记为 $[p_{ps_i},p_{vm_i},p_{os_i},p_{s_i},p_{m_i},p_{f_i},p_{db_i},p_{a_i}]$ 结合信息系统等级保护的定级 b_i,可计算每个实体的重要性(Importance I),记为: $I_{e_i}=wp_{e_i}+b_i$,其中 w 为权重.

行为操作:是指用户对数据所进行的操作,记为 bp_i $[OpName,OpType]$,其中,$OpName$ 代表操作的名称,包括查询、删除、修改和新增等,$OpType$ 代表操作类型,对应服务器、系统、数据库层面的操作,还包括

调整配置，备份等.

2　基于机器学习方法的用户行为分析模型

机器学习是指使用计算机从大量数据中找出潜在的规律，来提取数据中的知识. 适合从大量用户行为数据中，找出正常的用户行为具有的特征，用来判断其他的用户行为是否正常.

典型的机器学习任务包括回归、分类和结构化学习. 按照是否有经过标记的训练数据，可以把机器学习分为监督学习、半监督学习和非监督学习. 在完全没有训练数据时，根据外界环境对机器学习算法的输出结果提供“好”或者“不好”的反馈信息，让学习算法自动调整参数，输出更好的行动决策时，称为增强学习.

Google Brain 推出了机器学习开源环境 Tensorflow，可以使用 CPU 和 GPU 等计算资源，基于 Python 环境实现跨平台分布式计算，是当前机器学习开发环境中最好选择之一. Karas 是一个高级软件包，可以更加方便的调用 Tensorflow 内置的各种机器学习算法，极大的简化了机器学习的应用.

2.1　环境搭建

在企业私有云上，搭建 Hadoop 的分布式平台，使用 Flume 和 Kafka 采集并汇总所有信息系统、中间件、数据库和安全软件的日志，汇总并分类，统一存储到 HDFS，建立 Hive 数据仓库存储用户行为相关的数据，使用 Google 公司的 TensorFlow 机器学习开源软件平台和 Karas 整合工具，搭建基于机器学习的企业私有云用户行为分析框架，如图 1 所示.

图 1　私有云环境下基于机器学习的用户行为分析框架

Fig. 1　User behavior analysis framework based on machine learning in Private Cloud

2.2　用户行为特征选择

根据以上定义，可见一次用户行为涉及的实体有很多种，用以描述每个实体的属性有多个，称之为特征(Features) . 为了使用机器学习对用户行为进行分析，首先需要对用户行为的多个特征进行分类和选择，去处冗余、无意义和复杂的属性. 结合企业私有云信息系统集成整合的实际应用环境，选择有代表性的用户行为特征，分为行为时态、行为主体、行为客体、行为操作四类，如表 1 所示.

2.3　用户行为数据预处理

在使用机器学习分析用户行为之前，需要对选取的行为特征进行预处理，以便用数字的形式表示，作为机器学习的输入，进行计算和判定.

1) 统计数据预处理. 采集企业私有云环境下的物理服务器、虚拟机、操作系统、数据库、信息系统等环境实体的详细信息，存放到用户行为数据库中，形成用户行为环境实体表.

统计所有用户行为涉及到的实体 e_i 被访问的频度，计算其被访问的概率 p_{e_i}. 一个实体被访问的概率越大，它的重要性越高，记为 $I_{e_i}=f(p_{e_i})$，表示重要性是 p_{e_i}的函数. 根据企业私有云下各个实体之间的内在联系，用户对某个信息系统模块的访问行为将使行为客体对应的实体集内所有实体被访问的概率增加. 对大量行为记录的统计分析可以获得一个接近真实的相关实体被用户访问的概率分布.

2) 行为主体数据预处理. 一个用户有 UserName，AccountID，UserCreatedTime，System，UserClassification，UserRisk，UserTrustLevel，UserRole，BehaviorCount 等属性，描述用户行为的主体. 去除不相关和冗余的属性，找到最能够描述用户主体特征的特征子集进行分析.

对行为主体特征进行预处理的神经网络，输入层为 UserName，AccountID，CreatedTime，System 隐含层判断是否为正确的用户名，是否为合法账号，是否在

正确的时间创建,是否属于合法的信息系统,分别记为 RightName,LeagleID,LeagleTime 和 LeagleSystem,均为 Bool 型变量. 经过输出层的处理后,得到行为主体的合法性特征 L_{bs_i},也为布尔类型,取值为 0 为非法用户,取值为 1 为合法用户.

表1 用户行为的特征子集

Tab.1 Feature subsets of user behaviors

序号	属性分类	属性	描述	类型
1	行为时态 be_i	Ts_i	开始时间(*TimeStart*)	*DateTime*
2		Te_i	结束时间(*TimeEnd*)	*DateTime*
3		*IP*	用户 *IP* 地址(*IP Address*)	*string*
4		Loc_i	用户所在地址(*Location*)	*string*
5		*Terminal*	用户终端类型(*Terminal*)	*string*
6	行为主体 bo_i	*UserName*	用户名(*UserName*)	*string*
7		*AccountId*	账号编码(*AccountId*)	*long int*
8		*CreatedTime*	创建时间(*CreatedTime*)	*long int*
9		s_i	所属系统(*System*)	*string*
10	行为客体 bs_i	ps_i	物理服务器(*PhysicalServer*)	*int*
11		vm_i	虚拟机编号(*VirtualMachine*)	*int*
12		os_i	操作系统(*OperatingSystem*)	*int*
13		s_i	应用系统(*ApplicationSystem*)	*int*
14		m_i	应用系统模块(*Module*)	*int*
15		f_i	应用系统功能菜单(*Function*)	*int*
16		db_i	数据库(*DataBase*)	*int*
17		tb_i	数据库表(*Table*)	*int*
18		att_i	数据库属性列(*Attributes*)	*int*
19	行为操作 bp_i	*OpName*	操作名(*OpName*)	*string*
20		*OpType*	操作类型(*OpType*)	*string*

3) 行为客体数据预处理. 用户行为客体是各种信息系统、模块、功能及对应的数据库、表、记录或属性列,可以表示为形如 $D[d_1,d_2,\cdots,d_i,\cdots,d_n]$ 的数据集合,以标准的格式存储到数据库中,形成企业的数据仓库. 某一次用户行为涉及的企业数据集 $bo_i[ps_i,vm_i,os_i,s_i,m_i,f_i,db_i,tb_i,a_i]$,可用一个九元组表示,对每一个九元组,可根据访问频度求其重要性. 实体集的重要性可定义为每个实体重要性的加权平均值,记为 $I_{bo_i}=\left(\frac{1}{n}\sum_{1}^{n}(w_iI_{e_i}+b_i)\right)$,$w_i$ 表示实体 e_i 重要性 I_{e_i} 的权重,b_i 是其偏置,可以理解为等级保护级别. I_{bo_i} 取值可通过激活函数规范化到[0,1],根据某个人为设定的阈值 θ_1 判断行为客体实体集的重要与否,取值为0(不重要)或1(重要). 激活函数是一类非线性函数,可以把某个定义域的输入值经过变换,通过某些选定的阈值 θ,映射到0到1的连续区间,方便进行判定,常见的激活函数包括 sigmoid、tanh、relu 等.

行为客体预处理的神经网络输入层为用户行为涉及的实体集,以 x_i 表示该实体 e_i 被访问的概率,结合信息系统等级保护定级的等级结果,计算每个实体的重要性 I_{e_i},经过输出层处理后,输出实体集的重要性,为布尔类型数值,记为 I_{bo_i}.

4) 行为操作数据预处理. 根据用户行为记录中的操作序列 $O[o_1,o_2,\cdots,o_i,\cdots,o_n]$,可以评估每种操作的风险等级,记为 R_{o_i},一个操作序列的风险值取序列中的最高风险值 $R_{bp_i}=f[R_{o_1},R_{o_2},\cdots,R_{o_i},\cdots,R_{o_n}]=Max[R_{o_1},R_{o_2},\cdots,R_{o_i},\cdots,R_{o_n}]$,类似的,可使用 Sigmoid 激活函数规范化到[0,1],根据阈值 θ_2 判定用户操作序列是高风险还是低风险.

行为操作数据预处理的神经网络输入层是操作序列的操作名 *OpName*,根据操作的属性和作用,判断该操作序列是否对企业数据带来高风险,输出布尔型数值 R_{bp_i}.

5) 行为时态数据预处理. 用户行为的时态记为 $be_i[Ts_i,Te_i,Loc_i,Terminal]$,根据企业的信息安全管理规则 f 计算分析行为的时态特征 $N_{be_i}=f[Ts_i,Te_i,Loc_i,Terminal_i]$;代表用户操作时态是否正常,取值为0表示非正常时态,为1时,表示正常的时态.

行为时态预处理的神经网络,输入为用户行为发生的起止时间的差 $T_{e_i}-T_{s_i}$,用户使用的网络地址 *IP*,物理位置 *Location*,终端类型 *Terminal*,隐含层分别判断时间是否正常(*RightTime*),是否合法 *IP* 地址(*LeagleIP*),是否合法的物理位置(*LeagleLoc*)和是否为信任的终端(*TrustTerminal*),输出为用户行为时态是否正常(N_{be_i}).

至此完成了对用户行为的时态,主体、客体和操作进行预处理的过程,通过神经网络,输出了四个布

尔类型的特征值，实现了用户行为特征的提取.

2.4　建立用户行为分析的神经元

完成以上的用户行为特征分类和预处理，可以把用户行为 UB $[BE,BS,BO,BP]$ 的特征可以表示为: b_i $[N_{be_i},N_{bs_i},N_{bo_i},N_{bp_i}]$，建立一个用来处理用户行为的神经元，以 $I_{bo_i},N_{be_i},I_{bo_i},R_{bp_i}$ 作为神经元的输入，表示为输入向量为 x_i = $[x_1,x_2,x_3,x_4]$，记为 X. 设权值矩阵为 w_i = $[w_1,w_2,w_3,w_4]$，记为 W，偏置矩阵为 b_i = $[b_1,b_2,b_3,b_4]$，记为 B'，代表其他不可预知的参数影响，输出值为 Nbi，代表一个行为 b_i 是否正常，记为 Y，则函数集合 $Y = Nb_i = H(X) = WX + B'$ 表示了输入和输出之间对应关系的所有可能函数的集合. 用户行为分析的神经元模型如图 2 所示.

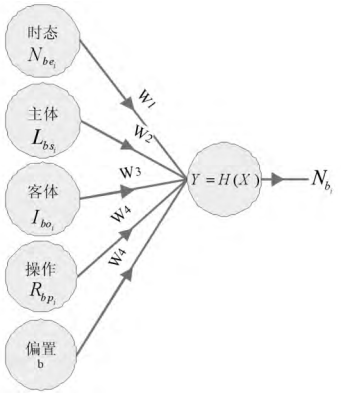

图 2　用户行为分析的神经元模型

Fig. 2　A neuron model for user behavior analysis

这个神经元可以进一步表示为 $Y = H(X) = sigmold(\sum_{i=1}^{4}(w_i x_i + b_i))$. 根据选定的阈值 θ_3，可以判断某个用户行为是否正常. 使用 Karas 可以用很少的代码调用 TensorFlow 建立图 2 所示的用户行为分析的神经元，与特征选择后的用户行为时态、主体、客体、操作特征预处理的神经元连接后，形成用户行为分析的神经网络. 使用训练数据对这个神经网络进行训练时，按比例切分训练数据，比如把 90% 的预处理后的用户行为记录作为训练数据，其余 10% 作为测试数据.

使用大量已标记的用户行为数据对神经网络进行训练，找到一个参数组 w_i,b_i，使得神经网络模型在训练数据集和测试数据集上的误差达到最小. 对应的函数即为目标函数，将用来判定后续的用户行为是否正常.

3　模型实现和效果分析

3.1　基于 TensorFlow 的神经网络实现.

在图 1 所示的用户行为机器学习框架中，在 Ubuntu，Mac OS，Win10 等操作系统上，运行 Python 开发环境，加载 TensorFlow 环境，并加载 Karas 工具，调用 Numpy 等数值计算开发包，使用内置的机器学习算法对用户行为分析机器学习模型进行训练、误差评估和参数优化，最终输出误差最小的机器学习模型进行用户行为自动分析.

建立神经网络、运行和训练的实现过程描述如下.

Step 1. 载入 TensorFlow 和 Numpy 环境;

Step 2. 从 TensorFlow 项目文件夹中载入用户行为数据文件;

Step 3. 定义神经网络图，建立数据预处理和行为分析的神经网络模型，并定义误差函数;

Step 4. 运行神经网络训练过程;

Step 5. 测试并调整参数，使得误差函数取值最小，输出最终模型用于判断后续用户行为.

3.2　用户行为分析神经网络预测效果

经过神经网络训练和优化，找到对训练数据和测试数据误差都能接近最小的神经网络参数，作为最终的神经网络输出. 能对后续产生的用户行为数据进行预测. 预测的过程同样经过数据采集、存储、初始化和神经网络判断，输出用户行为的结果. 在测试环境下，从信息系统日志中提取 12000 条用户行为数据，并把其中 10000 条满足信息新系统使用规范的数据标记为正常行为，其余标记为不正常行为，作为训练数据和测试数据，用来训练用户行为分析的神经网络. 根据不同的学习率(Learning Rate) ，采用基于梯度下降(Gradient Descent) 优化器来优化神经网络的参数，经过多次更新后，训练误差逐步下降. 图 3 显示了不同学习率时，经过一段时间的训练和参数更新，训练误差的变化曲线.

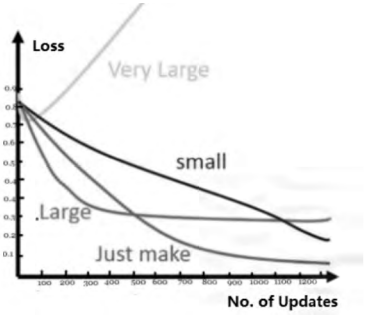

图 3　不同学习率时的误差曲线

Fig. 3　Loss curve for different learning rate

从图 3 中可见当学习率太大(Very Large) 时，模型不能找到合适的参数，导致训练误差迅速超出预期. 当学习率太小(Small) 时，误差下降得太慢，导致训练需要太长的时间. 学习率稍大(Large) 时，误差初期下降很快，而后期基本不会下降，导致训练迟迟不能完成. 当选择到刚好合适(Just Make) 的学习率时，机器学习算

法可以在适当的训练时间内达到满足需要的误差.

预测不准确的原因常常是因为输入的变量的选择过多,导致模型出现对训练数据的过拟合(Over Fitting).要减少过拟合,需要从几个方面入手,一是选择最有代表性的特征子集.二是修改特征提取和行为分析的函数集,使之更契合目标函数的变换关系.三是增加标记准确的训练数据,使模型调整参数后,能更加贴近真实的目标函数.

随着训练数据(Testing Data)的增加,机器学习算法能够逐步达到较低的训练误差和测试误差.误差变化随着训练数据增加的曲线如图4所示.

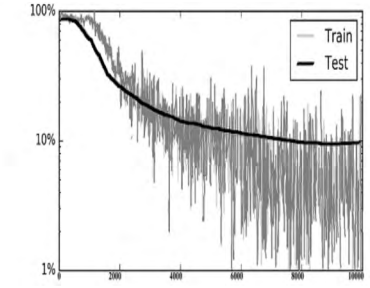

图4 不同数量的训练数据时的误差情况

Fig.4 Loss curve of different training data growth

考虑到机器学习过程中需要找到一个最合适的函数,使之无限逼近目标函数,而机器学习方法本身存在误差,给定的神经网络模型在不同的训练数据集和测试数据集上有不同的准确率,以少量预测案例的不准确,换取相同类型的用户行为的自动分析.在实际的应用中,用于判断用户行为的有监督神经网络模型在测试数据集上,预测准确率能达到90%,通过模型优化、参数调优和增加训练数据等技术手段,可以进一步提高,从而具有实际应用的可行性.

4 结束语

企业的信息系统在云环境下面临日益多元化的安全威胁,分析用户访问企业数据的行为可以帮助企业找到潜在的数据安全风险.本文在私有云环境下,提出了一个基于机器学习的用户行为分析模型,通过采集、汇总、存储大量信息系统日志数据,使用机器学习的方法自动分析海量用户行为,找出其中的异常行为,从而及时响应.考虑到长期性和持续性,数据安全的设计应该和信息系统设计同步进行,采集标准化的用户行为数据,予以准确标记,从而方便机器学习模型实现更高的预测精度,是未来的研究方向.

参 考 文 献

[1] Innovation Research Team. SIEM-Centric User Behavior Analytics (SCUBA) [EB/OL]. New Hotness in Security. Track3. http://trace3.com/wp-content/uploads /2016/09. 2016.

[2] 冯登国,张 敏,张 妍等.云计算安全研究[J].软件学报,2011,22(1):71-83.

[3] 房 梁,殷丽华,郭云川,等.基于属性的访问控制关键技术研究综述[J].计算机学报,2016,39(79):1-20.

[4] Younis A, Kifayat K, Merabti M. An access control model for cloud computing [J]. Journal of Information Security and Applications,2014.19(1): 45-60.

[5] 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.

[6] 陈亚睿.云计算环境下用户行为认证与安全控制研究[D].北京:北京科技大学,2011.

[7] 刘正南.云环境下基于用户行为评估的访问控制模型研究[D].西安:西北农林科技大学,2016.

[8] 冯登国,张 敏,李 昊.大数据安全与隐私保护[J].计算机学报,2014,37(1): 246-258.

[9] Lin G, Wang D, Bie Y, et al. MTBAC: A Mutual Trust Based Access Control Model in Cloud Computing [J]. China Communication, 2014. 11(4): 154-162.

[10] 刘 莎.Hadoop云平台的用户可信访问控制模型研究与实现[D].成都:四川师范大学,2014.

[11] 王电轻.基于hadoop的网站用户行为分析系统架构和实现[D].北京:中国科学院大学,2016.

[12] 何 苗.基于机器学习的移动数据安全检测技术研究[D].北京:北京邮电大学,2015.

[13] Chen S, Ghorbani M, Wang Y, et al. Trace-Based Analysis and Prediction of Cloud Computing User Behavior Using the Fractal Modeling Technique [C]//IEEE. Big Data (Big Data Congress) IEEE International Congress. Alaska: IEEE,2014: 733-739.

[14] Ghazinour K, Ghayoumi M. An autonomous model to enforce security policies based on user's behavior [C]// IEEE/ACIS. 14th International Conference on Computer and Information Science (ICIS). Las Vegas: IEEE Computer Society,2015: 95-99.

[15] Jaiganesh M, Aarthi M, Kumar A. A Fuzzy ART-based user behavior trust in cloud computing [J]. Advances in Intelligent Systems & Computing, 2015,(324): 341-348.

[16] Ladekar A, Pawar P, Raikar D, et al. Web Log based Analysis of User's Browsing Behavior [J]. International Journal of Computer Applications, 2015. 115(11): 5-8.