内部威胁检测中用户行为模式画像方法研究

郭渊博[1]，刘春辉[1,2]，孔菁[1]，王一丰[1]

（1. 中国人民解放军战略支援部队信息工程大学密码工程学院，河南 郑州 450001；2. 中国人民解放军 61213 部队，山西 临汾 041000）

摘 要：行为画像技术利用无标注历史数据构建用户行为“常态”，是检测企业内部威胁的有效手段。当前标签式画像方法依赖人工提取特征，多用简单统计方法处理数据，导致用户画像模型缺少细节、不够全面。提出了一种行为特征自动提取和局部全细节行为画像方法，以及一种行为序列划分和全局业务状态转移预测方法，能够较全面地刻画用户行为模式。构建了一个基于行为画像的内部威胁检测框架，将局部描写与全局预测相结合，提高了检测准确率。最后用 CMU-CERT 数据集进行了实验，AUC（area under curve）得分 0.88，F1 得分 0.925，可有效应用于内部威胁检测过程中。

关键词：行为序列；画像提取；内部威胁；隐马尔可夫模型

中图分类号：TP391

文献标识码：A

doi: 10.11959/j.issn.1000-436x.2018282

Study on user behavior profiling in insider threat detection

GUO Yuanbo[1], LIU Chunhui[1,2], KONG Jing[1], WANG Yifeng[1]

1. Cryptography Engineering Institute, Information Engineering University, Zhengzhou 450001, China

2. Unit 61213 of The Chinese People's Liberation Army, Linfen 041000, China

Abstract: Behavior profiling technic using no-labeled historical data to build normal behavior model is an effective way to detect insider attackers. The state-of-the-art labeled profile methods extract features artificially and process data by simple statistical methods, whose incomplete behavior model lacks details. An automated feature extracting and full-detail behavior profiling method as well as a behavior sequence splitting and business state transition predicting way was proposed. Combining above two methods, an insider threats detection framework was established, which improved detection accuracy. Experimenting with CMU-CERT data set, AUC (area under curve) score was 0.88 and F1 score was 0.925. With the better performance, it can be used in detecting insider threats.

Key words: behavior sequence, profiling extraction, insider threat, hidden Markov model

1 引言

全球企业每年因为内部用户蓄意破坏或无意失职而造成的损失所占比重越来越大，内部威胁日益成为企业安全关注的重点。2015 年美国网络犯罪调查显示，23%的电子犯罪事件来自于内部人员，45%的受访者认为内部人员攻击造成的损害要远高于外部攻击带来的损害。Verizon RISK Team 发布的《2017 年数据泄露调查报告》[1]指出，15%的数据泄露是由内部人员造成的。Crowd Research Partners 在 2018 年对 472 位资深网络安全专家进行的在线调查[2]显示，53%的组织确认过去一年内遭受过内部威胁攻击，29%的组织认为内部威胁攻击越来越频繁。

内部威胁攻击中，攻击者来自企业内部，攻击行为往往发生在工作时间，恶意行为嵌入大量正常数据中，增加了数据挖掘分析的难度。同时，内部攻击者往往具有组织安全防御机制的相关知识，会采取措施规避安全检测[3]。

收稿日期：2018–05–07；**修回日期**：2018–07–26

基金项目：国家自然科学基金资助项目（No.61602515，No.61501515）

Foundation Item: The National Natural Science Foundation of China (No.61602515, No.61501515)

用户在访问文件、使用应用程序、获取内部资源、使用设施设备的时间和频率等方面会形成一个相对固定的行为模式。相同角色、相同工作部门的用户的工作性质相近，其行为模式具有一定的相似性。若用户行为明显偏离正常模式，则表示该用户有意隐藏其恶意行为或存在违反企业相关政策（如知识产权政策）违规获取工作需求之外信息的行为。对用户正常行为进行画像，并加以对比分析，可以有效检测用户行为模式变化。

本文对内部威胁中用户行为模式画像方法进行了研究，针对标签式画像方法特征提取过度依赖人工提取，行为模型缺少细节、不够全面等问题，提出了一种全新的自动化行为细节特征提取方案，构建了细节描写与全局刻画相结合的用户行为模式画像框架。

用户对不同网站的访问模式、与某个 E-mail 账户的联系模式等，都可以用来描述该用户的历史性、习惯性行为。通过自动化提取行为细节特征，利用一分类支持向量机（OCSVM，one class support vector machine）集群构建全细节行为画像，可以判断用户行为是否与历史习惯存在明显差异。

用户每天进行大量业务操作，且其工作具有稳定性，业务流程具有固定性，因此用户每天的活动具有一定的重复性。采用隐马尔可夫模型整合多类行为数据，提取用户行为序列，可以揭示隐藏在行为背后的业务逻辑，预测业务流程的转移概率，刻画用户全局行为模式。根据转移概率的大小可以判定用户行为偏离历史行为的程度。

全细节行为画像与业务状态转移预测相结合的用户行为模式画像框架，能够充分提取并利用审计日志中的用户行为信息，较全面地刻画用户行为模式，有效提高企业内部用户异常行动判定的准确率。

2 研究现状

近年来，随着内部用户攻击行为对企业造成的影响越来越大，内部威胁检测也被越来越多的人关注。相关领域的专家、学者针对此类问题提出了不同的技术方法和解决方案。但由于内部攻击具有隐蔽性、多元性等特点[3]，内部威胁依然是企业组织面临的主要威胁之一。Nurse 等[4]整合并明确定义内部威胁的多方面因素，提出了表征内部用户攻击行为的框架，对表征用户心理状态变化的因素进行了研究。Legg 等[5]提出了基于树型结构的用户角色特征画像模型，人工定义了一系列描述用户日常行为的特征，并通过主成分分析（PCA，principal component analysis）方法对特征进行了降维，对每一个用户、每一个工作角色的活动记录进行特征提取，构建了刻画用户、角色行为的树型模型，然而，该方法特征提取过度依赖人工挑选，缺乏有效的自动化处理机制，无法与用户历史行为进行有效贴合。Rashid 等[6]首次将隐马尔可夫模型应用到内部威胁检测中，利用隐马尔可夫模型学习用户正常行为序列，通过对比发现明显偏离正常行为的用户，然而，单纯利用隐马尔可夫模型，并没有取得较高的检测准确率。Gamachchi 等[7]将图处理技术应用到内部威胁检测中，提出了图处理单元与异常检测单元相结合的理论框架，然而，只是将图理论用到了用户与设备、用户与行为的描述上，并没有关注用户行为的画像问题。Gavai 等[8]提出了利用企业在线活动和社交数据检测内部威胁的方法，利用非监督的独异森林方法检测统计意义上的异常点，准确率达到了 73.4%，利用监督学习方法，预测用户离职情况，取得了 0.77 的 AUC 分数，同样地，该方法没有关注用户行为画像问题。Parveen P[9]在系统中部署 k 个分类器，提出了一种基于“k-投票”形式的内部威胁解决方案，使用 Hadoop 分布式框架提高学习效率，然而，未针对内部威胁数据进行实验，无法判断其实际效果。文献[10]提出文件内容异常检测模型，该模型使用文本分割和朴素贝叶斯方法对企业内部文件内容进行分类，根据个体行为与群组行为偏移量检测文件访问异常行为，实验证明该模型对保护内部文件访问有一定作用，但只针对文件操作单域行为，且检测效果完全取决于所用词汇库的丰富程度。Ioannis 等[11]提出了活动树模型，记录用户的工作流模式，根据分支长度、对应节点相似度等指标判断新行为与历史工作流模式的匹配度。本文将正常用户行为日志存储至全文搜索引擎，通过搜索用户当前行为与历史行为的差异，形成行为特征向量，实现了特征提取的自动化，并能够有效衡量当前行为与历史行为的偏离程度。

由于内部威胁的复杂性和企业数据的隐私性，之前的研究多从某一个或几个维度对用户活动进行检测，存在检测准确率低、误报率高等问题。本研究结合用户历史行为提取特征，从单类行为细节和全局状态转移 2 个方面对用户行为进行综合画像，较全面地刻画了用户行为模式，异常行为检测

效果得到了明显提升。

3　思路和方法

在内部威胁检测中，由于攻击模式多样、攻击样本缺乏、人工标记标签工作量大等困难，当前较为成熟的有监督学习分类方法无法有效利用现有数据进行训练。用户行为画像技术无需标签数据，通过学习用户的历史行为模式，可以形成精细描绘用户行为的历史画像。本文的主要目标有：1) 研究行为特征自动提取和局部全细节行为画像方法、行为序列划分和全局业务状态转移预测方法；2) 将局部描写与全局预测相结合，搭建基于行为画像的内部威胁检测框架；3) 利用卡耐基梅隆大学计算机安全应急响应组（CMU-CERT, Carnegie Mellon University Computer Emergency Response Team）数据集对本文所述方法的有效性进行检验。本节主要介绍方法思路和涉及的理论基础。

3.1　审计日志获取和数据准备

随着企业安全意识的提高，安全策略、访问控制、权限管理等防护措施都基本完善，然而仅通过这些防护措施，并不能完全保证企业信息的安全。为保证合法用户有效访问受保护资源、防止非法用户非授权访问、保留用户行为记录进行违规追查，日志分析和审计成为保护企业信息安全、监控内部用户行为合规性的重要手段。在审计系统中，部署在企业内部的各类传感器会不断记录用户操作行为，并生成相关日志，存储至日志服务器。用户登录、文件操作、邮件收发、网页浏览、外设使用等行为，是企业审计用户行为所使用的最基本的数据，相比网络流量、能量消耗等数据，这 5 类数据具有采集方便、可理解性强的特点。

企业内部恶意活动往往以窃取信息、伪装身份、破坏系统为主要目的，其中发生最多的是窃取组织内部信息资产，这些信息资产包括但不限于用户数据信息、金融/财务信息、知识产权、内部人员信息等。信息窃取过程中，恶意员工通常通过企业的数据库服务器、文件服务器、OA 应用/业务应用、终端等获取内部信息，然后通过移动介质、邮件、网页上传等方式将信息转移出企业内部。身份伪装攻击多表现为恶意员工窃取合法员工的身份，冒充他人身份发布恶意信息、执行破坏性操作等。系统破坏攻击多为员工的报复性行为，通常表现为恶意删除关键数据、删除关键系统模块等行为。通过监控用户文件操作、邮件收发、外设使用、网络浏览行为可以防范信息窃取威胁，监控用户登录、邮件收发、文件操作可以防范身份伪装威胁，监控文件操作可以防范恶意系统破坏威胁。综上所述，综合分析员工审计日志中的登录行为、文件操作行为、邮件收发行为、外设使用情况、网页浏览记录等内容，提取用户的行为特征，进行行为画像，能够为检测内部威胁行为提供解决方法。

3.2　行为特征提取和全细节行为画像方法

企业审计日志中，用户行为本身没有分类标签，且很难及时准确地判断其是否具有威胁性。为海量日志人工标记标签不仅耗时耗力，而且无法保证准确性。

全文搜索引擎技术通过扫描文档中的每一个词，对每个词建立索引，指明该词在文档中出现的次数和位置，当用户查询时，检索程序就根据事先建立的索引进行查找，并将查找的结果即时反馈给用户[12]。

为用户的历史行为日志建立索引，并存储到搜索引擎数据库中。当新的行为数据到来时，检索该行为模式在历史行为中出现的次数以及出现的时间节点等信息，通过与历史行为对比，可以判断新行为是否为异常操作。本文将全文搜索引擎技术作为联系用户新行为和历史行为的桥梁。在训练阶段，将之前某一段时间内的历史用户行为数据作为正常数据索引并存储至搜索引擎数据库中，作为初始搜索的基础数据。之后，对新的行为进行全文搜索，可以得到新行为是否出现在历史行为记录中以及在历史行为中的占比，进而将字符型的日志数据转化为方便处理的数值型向量。

由于用户历史行为不具有分类标签，传统的二分类方法不能很好地适应该问题。在模型训练阶段，本文假设开始一段时间内用户行为日志中不包含恶意行为。由于一分类支持向量机对数值型向量分类具有良好的分类效果，本文将 OCSVM 作为基础分类器。OCSVM 集群能够有效降低单模型中数据过拟合导致的误报、漏报问题带来的影响，并能够随着时间推移学习用户行为模式变化，实现行为模式的在线更新，提高建模的健壮性和稳定性，于是，本文提出利用一分类支持向量机集群对用户历史行为模式进行细节画像的方法。

首先将用户单类行为序列按时间顺序，以某固定时间窗口（例如 7 d）为单位划分为不同的行为

块。同时，保证每一个行为块中包含用户工作日和休息日的行为数据，这样可以较为全面地描述一段时间内的行为模式。利用每一个行为块中的数据训练得到一个 OCSVM 分类器。保存时间最近的 v 个数据块形成的分类器集合 $M=\{M_1,M_2,\cdots,M_v\}$，构成 OCSVM 集群。当新来数据时，取 M 个分类器得分的平均值作为新数据的异常得分。一分类支持向量机集群如图 1 所示。

3.3 隐马尔可夫全局画像和行为序列划分方法

一分类支持向量机集群可以对用户单个行为的细节进行画像，能够有效判断单类行为的异常。但是当内部攻击者具有组织安全防御机制的相关知识，并采取一定的规避措施时，仅利用单类行为判断用户是否存在攻击行为的准确性有所降低。此时，整合多类行为数据，能够更好地刻画用户全局行为模式。本文采用隐马尔可夫模型，提取用户全局行为序列，揭示隐藏在行为背后的业务逻辑，预测业务状态的转移概率。

隐马尔可夫模型（HMM，hidden Markov model）是结构最简单的动态贝叶斯网络，是一种著名的有向图模型，主要用于时序数据建模[13-14]。隐马尔可夫模型是马尔可夫链的一种，它的状态不能直接地观察到，但能通过观测向量序列观察到，每个观测向量都是通过某些概率密度分布表现为各种状态，每一个观测向量由一个具有相应概率密度分布的状态序列产生[14]。

如图 2 所示，隐马尔可夫模型中的变量可以分为 2 组。第一组是状态变量$\{y_1,y_2,\cdots,y_n\}$，其中 $y_i \in Y$ 表示第 i 时刻的系统状态。通常假定状态变量是隐藏的、不可被观测到的，因此状态变量又称为隐变量。第 2 组是观测变量$\{x_1,x_2,\cdots,x_m\}$，其中 $x_i \in X$ 表示第 i 时刻的观测值。在隐马尔可夫模型中，系统通常在多个状态$\{s_1,s_2,\cdots,s_N\}$之间转换，因此状态变量y_i的取值范围Y通常是有N个可能取值的离散空间。

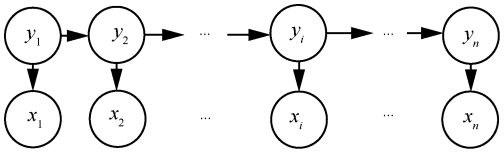

图 2 隐马尔可夫模型的图结构

图 2 中的箭头表示了变量间的依赖关系。在任一时刻，观测变量的取值仅依赖于相对应的状态变量，与其他状态变量及观测变量的取值无关。同时 t 时刻的状态 y_t 仅依赖于 t–1 时刻的状态 y_{t-1}，与其余状态无关。基于这种依赖关系，所有变量的联合概率分布为

$$P(x_1,y_1,\cdots,x_n,y_n)=P(y_1)P(x_1\mid y_1)\prod_{i=2}^{n}P(y_i\mid y_{i-1})P(x_i\mid y_i) \tag{1}$$

状态转移概率是指模型在各个状态间转移的概率，通常记为 $A=[a_{ij}]_{N\times N}$。其中 $a_{ij}=P(y_{t+1}=s_j\mid y_t=s_i),1\leqslant i,j\leqslant N$，表示在任意时刻 t 若状态为 s_i，则在下一个时刻状态为 s_j 的概率。

企业安全审计系统中，由于不同活动的监控器传感器不同，同一类活动中所有用户数据混合在一起。为方便后续操作，需要将行为日志进行预处理。首先，将日志数据库中的不同日志数据，按照用户 ID 进行重新划分，将每个用户的全部行为数据放到一个独立的文件中。随后，将每个用户的行为按照发生的时间顺序进行排序，得到用户行为数据流。

在隐马尔可夫模型中，要求观测序列离散且有限。因此需要将用户行为数据流中的长序列划分为

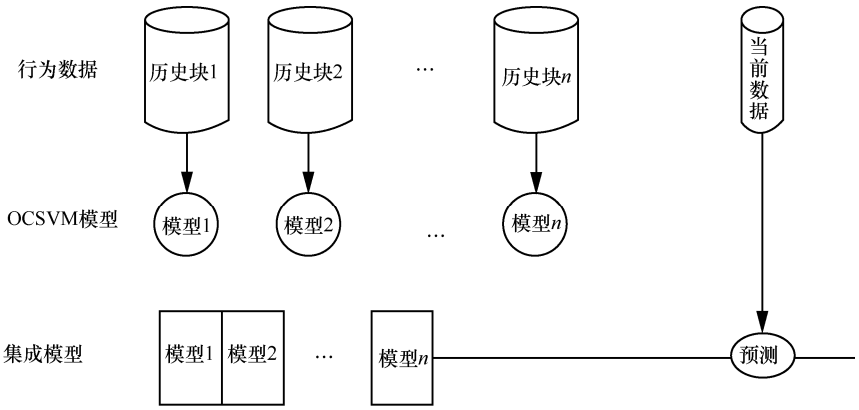

图 1 一分类支持向量机集群

便于处理的短序列。在现实生活中，由于用户处理的业务流程不同，产生的行为序列也会存在差异。在进行业务切换时，用户行为的间隔时间比业务进行时的间隔时间长。基于以上情况，本文假设相同业务状态中用户行为间隔时间 Δ 小于时间间隔阈值 θ，在业务状态发生切换时，$\Delta>\theta$。根据时间间隔阈值 θ 可以将用户行为序列流划分为多个具有先后顺序的短序列。

在相同的业务流程中，用户的行为序列应大致相同。为保证观测值的有限性，提升 HMM 模型预测效率，将划分好的短序列根据莱文斯坦比进行 k-means 聚类。相似的短序列被聚到同一类中，于是，在进行 HMM 模型训练时，可以用类名称代替该类中的所有短序列。得到观测值集合 $X=\{x_1,x_2,\cdots,x_m\}$，其中 $x_i \in X$ 表示第 i 时刻的观测值所在的类名称。

4　系统实现

在内部威胁检测中，由于获取攻击行为样本代价高、难度大，且正负例样本比例严重失衡，传统的二分类方法不能很好地适应该问题。因此，在画像提取部分，只能对单类行为细节和全局行为序列特征进行学习，并分别形成一个对正常用户行为的数据描述模型。而后，根据设定的阈值判断新行为样本的归属。利用上文介绍的隐马尔可夫模型和单分类支持向量机集群构建集成学习方法，组成一个提取用户画像、计算用户行为异常得分的框架。用户行为模式画像框架如图 3 所示。

4.1　日志解析器

由于企业数据的机密性、隐私性等原因，目前无法获取到真实企业中的数据为本文方法进行训练和测试。因此，使用目前认可度较高的 CMU-CERT 集成数据集作为实验数据源。

CMU-CERT 数据集是由美国国防部高级研究计划局（DARPA, defense advanced research projects agency）赞助的卡耐基梅隆大学内部威胁研究中心与 ExactData 公司合作，从真实企业环境中采集数据构造的一个内部威胁测试集。该数据集模拟了恶意内部用户实施系统破坏、信息窃取与身份伪装 3 类主要的攻击行为。除攻击行为数据外，还包含了大量正常的背景数据。在该数据集中，企业审计日志包含 5 个分别记录不同用户活动的文件。这 5 类活动是登录（login）、外设使用（device）、电子邮件（e-mail）、网页（Web）、文件读写（file access）。解析每一条数据可以得到时间戳（timestamp）、用户 ID（userID）、设备 ID（deviceID）、活动名称（activity）等信息，部分活动可能包含更多的信息，本文中统称为活动属性（attribute），例如电子邮件包含收件人、发件人、邮件内容等。CMU-CERT 数据集用户行为活动的具体内容如表 1 所示。

表 1　CMU-CERT 数据集用户行为活动内容

类别	包含活动	数量
登录	登入（login）、登出（logoff）	3 530 286
外设	连接（connect）、断开（disconnect）	1 551 829
邮件	发送（send）、浏览（view）	10 994 958
网页	访问（visit）、上传（upload）、下载（download）	117 025 217
文件	读（read）、写（write）、删除（delete）、复制（copy）	2 014 884

针对不同活动的属性，在解析过程中，需要进行一定的处理。在电子邮件活动中，考虑到真实企业环境中邮件内容的机密性，将邮件内容和附件信息直接舍弃，不进行处理；在发送的邮件中，将收件人信息加入活动属性；在接收的邮件中，将发件人信息加入活动属性。在文件读写活动中，将路径和文件名加入活动属性。在网页浏览中，将 URL 信息加入活动属性。登录和外设使用两类活动不包含属性数据，将其活动属性设为空(None)。

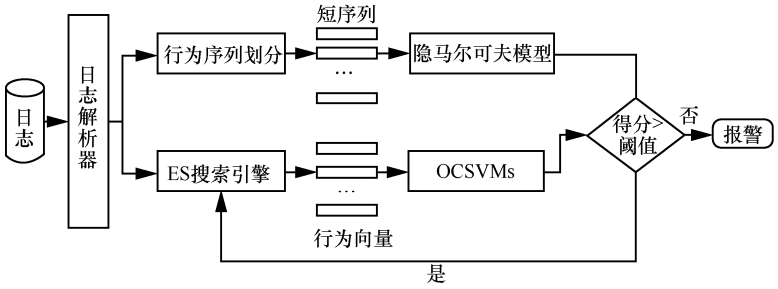

图 3　用户行为模式画像框架

最终，每一条日志可以解析为一个五元组(*timestamp*, *userID*, *deviceID*, *activity*, *attribute*)。

4.2 行为序列划分和全文搜索引擎特征提取

图 3 中，日志解析器首先将原始日志数据按照用户 ID 划分为不同的数据流。随后将数据流分别进行行为序列划分和特征提取。行为序列划分如算法 1 所示。

算法 1 用户行为状态划分

输入 用户行为序列 $\boldsymbol{A}=(a_1,a_2,\cdots,a_i,\cdots)$

输出 行为观测状态$\{x_1,x_2,\cdots,x_n\}$

1) $\boldsymbol{\theta}$←时间间隔阈值

2)***first_index***←1

3) ***last_index***←1

4) ***s_index***←1

5) for *i* in (*index* of ***A)*** do

6) t_i←time($\boldsymbol{a}_2$) - time($\boldsymbol{a}_1$)

7) if $t_i>\boldsymbol{\theta}$ do

8) ***last_index***←*i*

9) $serial_{s\_index}$← ($a_{first\_index},\cdots,a_{last\_index}$)

10) ***s_index***←***s_index*** + 1

11) ***first_index***←*i*

12) end if

13) end **for**

14) 聚类集合 $X=\{x_1,x_2,\cdots,x_n\}$←将 *serial* 按相似度聚类

15) return *X*

算法 1 首先计算相邻行为之间的时间间隔，当时间间隔 t_i 小于时间间隔阈值 $\boldsymbol{\theta}$ 时，将行为划分到相同的短序列中；当时间间隔 t_i 大于时间间隔阈值 $\boldsymbol{\theta}$ 时，开启一个新序列，后一个行为被划分到新序列中。当用户行为序列按照时间间隔阈值划分为多个短序列时，根据序列相似度，对所有短序列进行聚类。最后，输出聚类后的序列集合。

ES（elastic search）是一款基于 apache lucence 的开源的实时分布式搜索和分析引擎，能够以极高的速度处理大规模数据，实现稳定、可靠、快速地实时搜索，是当前流行的企业级搜索引擎。在实现过程中，本文使用开源的 ES 作为全文搜索引擎的技术支撑。

对于每条活动记录，按算法 2 进行全文搜索，得到特征向量 ***V***。

算法 2 中，对 *userID*, *deviceID*, *activity*, *attribute*， *timestamp* 进行组合查询，得到相应的查询数 *hit_num*，并计算不同 *hit_num* 之间的比值关系，最终输出由不同的比值关系组成的行为向量。算法 2 中，用户行为时间区间设置为±30 min，实现过程中可以根据实际效果对该值进行调整。

算法 2 基于全文搜索引擎的单类行为特征提取方法

输入 字符型行为 5 元组(***timestamp***, ***userID***, ***deviceID***, ***activity***, ***attribute***)

输出 数值型行为向量[*pc_per_user*, *act_per_user*, *act_time_per_user*, *attri_per_user*_1, *attri_per_user*_ 2, *attri_per_user*_3 ...]

1) *total_userID*←在 ES 中搜索 ***userID***，得到 *hit_num*

2) *user_pc_num*←在 ES 中搜索 ***userID& deviceID***，得到 *hit_num*

3) *pc_per_user*←*user_pc_num/ total_userID*

4) *user_act_num*←在 ES 中搜索 ***userID& activity***，得到 *hit_num*

5) *act_per_user*←*user_act_num / total_userID*

6) *user_act_timerange_num* 在ES中搜索 ***userID & activity&*[*timestamp*−30min, *timestamp*+ 30min]**，得到 *hit_num*

7) *act_time_per_user*←*user_act_timerange_num/ total_userid*

8) for *attritmp* in ***attribute*** **do**

9) *attri_act_user_num*←在 ES 中搜索 ***userid & activity&attritmp***，得到 *hit_num*

10) *attri_per_user_i*←*attri_act_user_num/ total_userid*

11) end **for**

12) return [*pc_per_user*, *act_per_user*, *act_time_per_user*, *attri_per_user*_1, *attri_per_user*_2, *attri_per_user*_3 ...]

从算法 2 中可以看出，*pc_per_user* 值代表用户使用此台设备的频率，当某用户在一台不常用的设备上进行操作时，该值趋近于 0；若某用户在常用设备上进行操作时，该值趋近于 1。同样地，当用户执行的活动与其惯有活动存在明显差异时，*act_per_user* 也趋近于 0。当用户在某个异常时间点进行惯有操作时，*act_time_per_user* 趋近于 0。当用户访问了历史中不常访问网页或者拷贝了某项机密文件时，*attri_per_user_i* 趋近于 0。综合上述分析，当用户操作出现异常时，特征向量 *V* 中的一

项或几项值会趋近于 0。

4.3　行为画像异常得分计算

前序活动完成后，将经过处理的行为类别和行为向量输入行为画像器进行异常得分计算。行为画像器中异常得分的计算过程如图 4 所示。

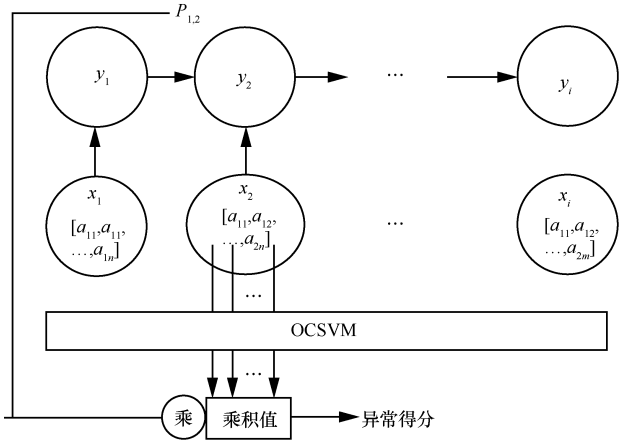

图 4　异常得分计算过程

图 4 中行为序列$\{a_{11},a_{12},\cdots,a_{1n}\}$构成观测状态 x_1，x_1 对应隐状态 y_1。P_{12} 表示隐状态 y_1 向 y_2 转移的概率。当新序列$\{a_{21},a_{22},\cdots,a_{2m}\}$到来时，可以得到计算 $P_{1,2}$。对于每一个行为 a_{2i}，利用之前训练好的 OCSVM 集群可以得出该行为的异常得分 s_i。最终，新序列的异常得分为

$$S = P_{1,2}\prod_{i=1}^{m} s_i \tag{2}$$

当隐状态 y_1 向 y_2 转移的概率很大，且行为序列中的活动为用户历史常做活动时，异常得分 S 值趋近于 1。反之，当隐状态 y_1 向 y_2 转移的概率很小，或行为序列中出现历史罕见活动时，S 值趋近于 0。

最后，根据选定的得分阈值判定当前行为是否为异常。判定为异常行为时，系统向安全运维人员发出警报；判定为正常行为时，将当前行为数据存储至全文搜索引擎，更新历史用户行为模式。

5　实验验证

为验证文中所提方法的有效性，本文利用 Python 语言开发了原型测试系统。测试环境操作系统为 CentOS 7，CPU 为 Intel i7-4790 @3.60 GHz，RAM 为 16 GB，硬盘为 1TB 机械硬盘。

5.1　实验数据

CMU-CERT 数据集包含企业内部 4 000 名用户 500 d 的所有活动记录，部分记录为攻击活动。由于实验条件的限制，对所有用户数据进行实验，会耗费大量的时间，且不利于模型参数的调校。本文选取用户 CMP2946 和用户 CDE1846 进行实验验证。根据 CMU-CERT 中的异常行为标签，可知该用户前 90 d 的数据中不包含攻击数据。实验中选取前 90 d 的数据作为训练数据，提取正常用户行为画像，剩下 410 d 的数据作为测试数据。测试过程中，当用户行为判定为正常后，也会被加入到正常数据中，用于充实和更新正常用户画像。

根据 CERT 数据集的介绍，用户 CMP2946 和用户 CDE1846，涉及两个完全不同的内部威胁场景。用户 CMP2946 从某天开始浏览求职网站，向竞争单位发出求职申请，并在离开公司之前，使用可移动存储设备偷窃公司数据。用户 CDE1846 登录其他用户的机器找寻机密文件信息，将找到的机密文件通过 E-mail 发送到私人邮箱中。两用户所有活动中包含的恶意行为如表 2 所示。

表 2　　内部威胁场景中恶意行为信息

用户 ID	场景	涉及记录	天数/d	恶意活动
CMP2946	文件窃取	242	29	13 E-mail、70 device、159 http
CDE1846	身份盗用	134	9	70 E-mail、46 file、18 login

5.2　评价标准

为评判本文提出方法的有效性，需要结合多个不同的指标作为评价标准。在内部威胁检测中仅用一个指标很难准确评价系统的效果，查准率（P，precision）、查全率（R，recall）、F1 得分（F1，F1-score）是经常用来评价模型效果的重要指标。

对于二分类问题，可将样例根据其真实类别与分类器预测类别的组合划分为真正例（TP，ture positive）、假正例（FP，false positive）、真反例（TN，true negative）、假反例（FN，false negative），令 *TP*、*FP*、*TN*、*FN* 分别表示其对应的样例数，则显然有 *TP*+*FP*+*TN*+*FN*=样例总数。分类结果的混淆矩阵如表 3 所示。

表 3　　分类结果混淆矩阵

真实情况	预测结果	
	正例	反例
正例	*TP*（真正例）	*FP*（假正例）
反例	*TN*（真反例）	*FN*（假反例）

查准率 P 和查全率 R 是一对矛盾的度量，一般来说，查准率高时，查全率往往偏低，反之亦然。查准率 P、查全率 R 分别定义为

$$P=\frac{TP}{TP+FP} \tag{3}$$

$$R=\frac{TP}{TP+FN} \tag{4}$$

F_1 是基于查准率和查全率的调和评价，定义如式(5)所示。

$$F_1=2\times\frac{PR}{P+R} \tag{5}$$

受试者工作特性曲线（ROC 曲线，receiver operator characteristic curve）是反映敏感度和特异度连续变量，评价系统有效性的综合型指标。根据学习器的预测结果对样例进行排序，按此顺序逐个把样本作为正例进行预测，每次计算出假正例率和真正例率，分别以它们作为横、纵坐标作图就得到了 ROC 曲线。AUC(area under curve)即 ROC 曲线下面的面积，AUC 越大模型效果越好。一般来说，AUC 接近 1 时，实验取得了较理想的效果，AUC 在 0.7~0.9 时，实验的准确性较高。

5.3　实验结果

在划分观测序列时，不同的时间间隔对序列长短、序列数量以及序列划分的有效性会产生不同的影响。为选取合理的时间间隔，方便后续实验开展，在其他实验参数固定的情况下，分别对 2 min、5 min、10 min 的时间间隔进行验证。图 5 展示了 2 个不同用户在不同的时间间隔下，14 d 活动的异常得分分布情况。可以看出，当时间间隔为 2 min 时，序列划分数量比 5 min、10 min 时要多，且得分分布较分散，此时异常得分均值分别为 0.37 和 0.205，方差分别为 0.073 和 0.037。时间间隔为 10 min 时，序列数量急剧减少，一个序列中包含的活动数量增加，但当一天中用户活动数量较少时，不能很好地表现用户工作状态的转换情况，此时得分的均值分别为 0.228 和 0.069，方差分别为 0.039 和 0.008。综上比较，最终选择 5 min 为合理的活动序列划分时间间隔，此时得分的均值分别为 0.217 和 0.151，方差分别为 0.027 和 0.034。

将 5 min 作为用户行为观测序列的划分间隔，进行后续实验。用训练好的画像器预测剩下的 410 d 中的活动，得到图 6 中的异常得分图。从图 6 可以看出，随着时间的推移，每个活动的异常得分趋于

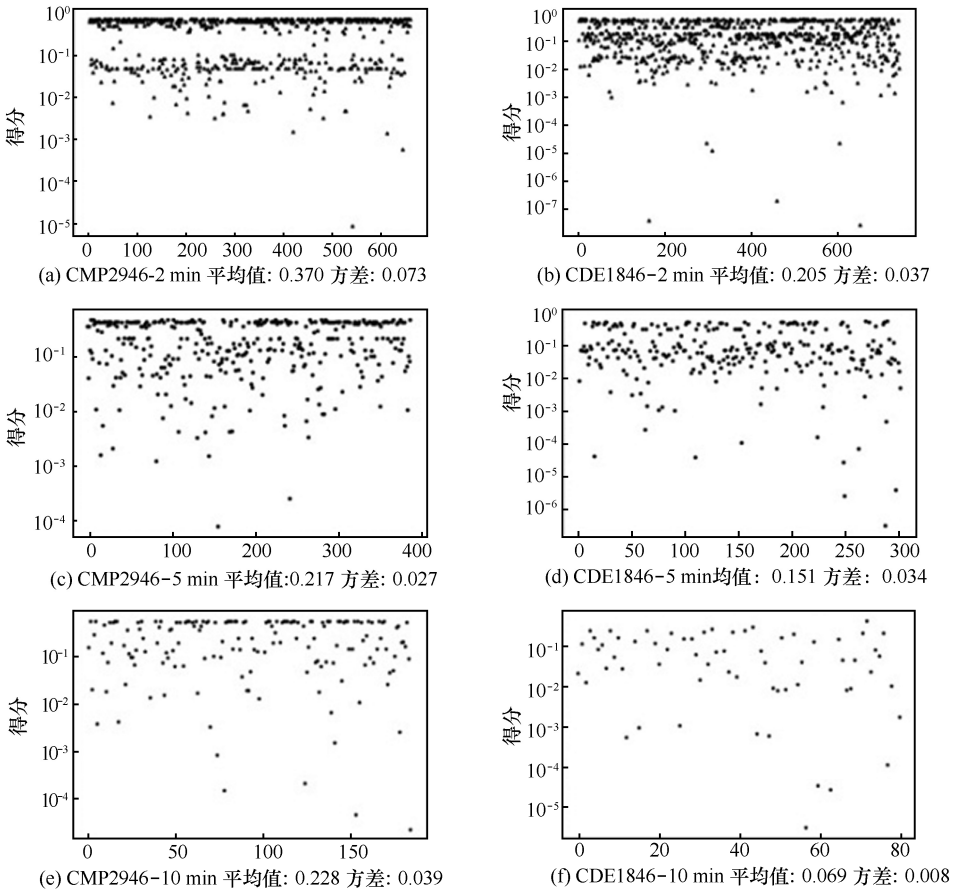

图 5　不同时间间隔下异常得分分布

平稳，由于隐马尔可夫模型中部分隐状态的转换概率较小，存在少部分活动得分小于1×10^{-4}。

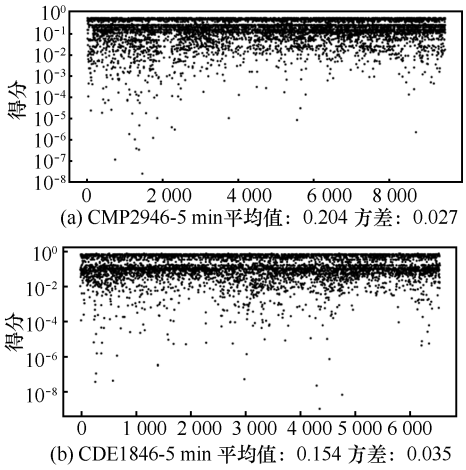

图 6　测试数据异常得分

选取不同的得分阈值作为异常行为的评判标准，训练模型的检测效果也会有所不同，如表 4 所示。

在实际工作中，不同的企业对查全率和查准率的要求不同。在安全级别较高的企业中，漏报恶意行为会引起较大损失，该类企业更倾向于低漏报率。一般企业中，当异常行为排查难度大时，高误报会加大安全工程师的工作量，降低正常员工的满意度和工作积极性，该类企业更倾向于低误报率。企业可以根据自身特征选择合适的评定阈值。从表 4 中，可以得出，用户 CMP2946、用户 CDE1846、整体均在 10^{-7} 时取得最大 *F*1。通过实验结果可以看出，该系统中，异常行为的得分趋近于 0。

系统的 ROC 曲线如图 7 所示。从图 7 中可以看到用户 CMP2946 的 *AUC* 达到了 0.95，系统整体的 *AUC* 为 0.88。

为充分表现本文所提方法的有效性，就查准率 *P*、查全率 *R*、*F*1 得分、*AUC* 这 4 项指标与现有研究方法进行了对比。本文第 2 节对内部威胁检测领域中现有方法进行了介绍，并对部分方法的优缺点进行了分析。鉴于上述文献中的方法并未提供用于测试的开源代码，且根据文献描述无法完全复现实验细节，这里选取与本文使用相同数据集(CMU-CERT)进行实验的文献[5]和文献[6]作为性能对比对象。文献[5]中使用查准率 *P*、查全率 *R* 这 2 项指标对实验结果进行的评价，此处根据 *F*1 定义计算得出其 *F*1 得分。文献[6]中使用 *AUC* 作为评价实验结果的标准，但全文未出现查准率 *P*、查全率 *R* 的确切数值。表 5 展示了 3 种方法的性能对比，表中文献[5]和文献[6]的数据选取自其原文中最优的实验结果。

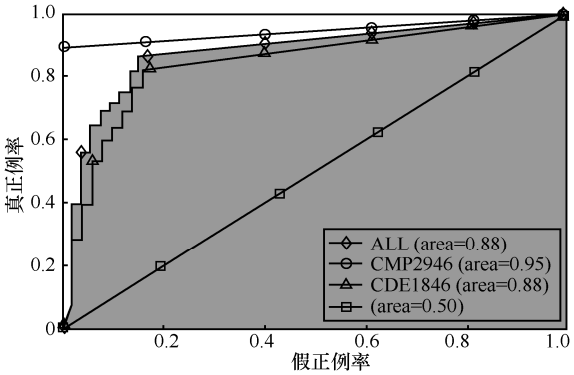

图 7　系统 ROC 曲线

表 5　　实验结果对比

指标	本文结果	文献[6]	文献[7]
查准率 *P*	0.999	0.420	—
查全率 *R*	0.860	1.000	—
*F*1 得分	0.925	0.591	—
AUC	0.880	—	0.830

从表 5 中的数据对比可以看出，本文方法在保证高查准率的同时，能够得到较高的查全率，*F*1 得分为 0.925 远高于文献[5]的 0.591。同时，本文方法整体 *AUC* 得分为 0.88，大于文献[6]中最优实

表 4　　不同得分阈值下模型效果

阈值	CMP2946			CDE1846			整体		
	P	*R*	*F*1	*P*	*R*	*F*1	*P*	*R*	*F*1
10×10^{-2}	1.000	0.846	0.916	0.999	0.746	0.854	0.999	0.805	0.892
10×10^{-3}	1.000	0.878	0.935	0.998	0.796	0.886	0.999	0.845	0.916
10×10^{-4}	1.000	0.885	0.939	0.998	0.812	0.895	0.999	0.855	0.922
10×10^{-5}	1.000	0.887	0.940	0.998	0.816	0.898	0.999	0.858	0.923
10×10^{-6}	1.000	0.888	0.941	0.998	0.819	0.900	0.999	0.860	0.9224
10×10^{-7}	1.000	0.888	0.941	0.998	0.819	0.900	0.999	0.860	0.925

验结果 0.83。通过对比，进一步证明本文提出的方法具有可行性。

6 结束语

随着信息化时代的全面到来，企业核心业务及机密信息都存储于信息系统。内部威胁攻击发生在企业内部，具有隐蔽性强、破坏性大的特点，直接威胁到企业的核心利益，造成严重危害。

本文针对当前标签式画像方法特征提取过度依赖人工，对用户行为模式画像缺少细节、不够全面等问题，提出了一种使用全文搜索方法的全新的自动化行为细节特征提取方案。通过自动化提取行为细节特征，利用一分类支持向量机集群构建了全细节单类行为画像。采用隐马尔可夫模型，整合多类行为数据，提取用户行为序列，揭示隐藏在行为背后的业务逻辑，预测业务流程的转移概率，刻画了用户全局行为模式。通过构建全细节行为画像与业务状态转移预测相结合的用户行为模式画像框架，充分提取并利用了审计日志中的用户行为信息，可以较全面地刻画用户行为模式，有效提高企业内部用户异常行动判定准确率。

利用 CMU-CERT 数据集对方法的有效性进行了验证，异常行为检测查准率为 0.999，单用户的 AUC 得分高达 0.95，系统整体 AUC 得分为 0.88，充分证明了本文方法的有效性。本文提出的框架为企业安全运维人员建立用户行为模式画像，有效检测恶意用户行为具有一定的借鉴意义。

参考文献：

[1] BAKER W, HYLENDER A, PAMULA C D, et al. 2017 data breach investigations report[R]. Verizon RISK Team, 2017: 49.

[2] SCULZE H. Insider threat spotlight report 2018[R]. Crowd Research Partners, 2018.

[3] 杨光, 马建刚, 于爱民, 等. 内部威胁检测研究[J]. 信息安全学报, 2016(3): 21-36.
YANG G, MA J G, YU A M, et al. Survey of insider threat detection[J]. Journal of Cyber Security, 2016(3): 21-36.

[4] NURSE J R C, BUCKLEY O, LEGG P A, et al. Understanding insider threat: A framework for characterizing attacks[C]//Security and Privacy Workshops (SPW), 2014: 214-228.

[5] LEGG P A, BUCKLEY O, GOLDSMITH M, et al. Automated insider threat detection system using user and role-based profile assessment[J]. IEEE Systems Journal, 2015.

[6] RASHID T, AGRAFIOTIS I, NURSE J R. A new take on detecting insider threats: exploring the use of hidden markov models[C]//The 2016 International Workshop on Managing Insider Security Threats. 2016: 47-56.

[7] GAMACHCHI A, SUN L, BOZTAS S. Graph based framework for malicious insider threat detection[C]//The 50th Hawaii International Conference on System Science. 2017: 2638-2647.

[8] GAVAI G, SRICHARAN K, GUNNING D, et al. Supervised and unsupervised methods to detect insider threat from enterprise social and online activity data[J]. JOWUA, 2015, 6(4): 47-63.

[9] PARVEEN P. Evolving insider threat detection using stream analytics and big data[M]. The University of Texas at Dallas, 2013.

[10] LIU A, MARTIN C, HETHERINGTON T, et al. A comparison of system call feature representations for insider threat detection[C]//Information Assurance Workshop, Proceedings from the Sixth Annual IEEE SMC.2005: 340-347.

[11] AGRAFIOTIS I, LEGG P A, GOLDSMITH M, et al. Towards a user and role-based sequential behavioral analysis tool for insider threat detection[J]. J. Internet Serv. Inf. Secur., 2014, 4(4): 127-137.

[12] 周敬才, 胡华平, 岳虹. 基于 Lucene 全文检索系统的设计与实现[J]. 计算机工程与科学, 2015, 37(2):252-256.
ZHOU J C, HU H P, YUE H. Design and implementation of Lucene-based full-text retrieval system[J]. Computer Engineering and Science, 2015, 37(2):252-256.

[13] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
ZHOU Z H. Machine Learning[M]. Beijing: Tsinghua university press, 2016.

[14] EDDY S R. Hidden Markov models[J]. Current Opinion in Structural biology, 1996, 6(3): 361-365.

[作者简介]

郭渊博（1975−），男，陕西周至人，博士，中国人民解放军战略支援部队信息工程大学教授、博士生导师，主要研究方向为大数据安全、态势感知。

刘春辉（1990−），男，山东安丘人，解放军 61213 部队助理工程师，中国人民解放军战略支援部队信息工程大学硕士生，主要研究方向为网络安全、用户画像。

孔菁（1993−），女，辽宁营口人，中国人民解放军战略支援部队信息工程大学硕士生，主要研究方向为网络安全、异常检测。

王一丰（1994−），男，江苏泰兴人，中国人民解放军战略支援部队信息工程大学硕士生，主要研究方向为多步网络安全、深度学习。