基于用户画像的自适应内部威胁检测模型*

李俊强[1],黄 洪[1,2],周子云[1]
(1.四川轻化工大学 计算机科学与工程学院,四川 宜宾 644005;
2.企业信息化与物联网测控技术四川省高校重点实验室,四川 宜宾 644005)

摘 要:内部威胁领域中,员工是内部威胁事件发生的主要研究对象。针对内部威胁检测系统中员工行为数据利用不全及检测细粒度低的问题,提出将用户画像作为内部威胁检测的基础,实现了员工行为信息的全方位构建并提高了检测的细粒度。通过引入滑动窗口机制对每个员工的画像模型进行自适应偏移,使得内部威胁检测模型能够实时检测威胁员工。检测模型在 CERT4.2 数据集中进行验证,取得较好结果。

关键词:内部威胁检测;用户画像;自适应
中图分类号:TP309.2 文献标识码:A DOI:10.19358/j.issn.2097-1788.2022.05.007
引用格式:李俊强,黄洪,周子云. 基于用户画像的自适应内部威胁检测模型[J].网络安全与数据治理,2022,41(5):43-48.

An adaptive insider threat detection model based on user portrait

Li Junqiang[1], Huang Hong[1,2], Zhou Ziyun[1]
(1.School of Computer Science and Engineering, Sichuan University of Science & Engineering, Yibin 644005, China;
2.Key Laboratory of Higher Education of Sichuan Province for Enterprise Informationalization and Internet of Things, Yibin 644005, China)

Abstract: In the field of internal threat, employees are the main research objects of internal threat events. To solve the problem of incomplete utilization of employee behavior data and low fine granularity in the insider threat detection system, this paper proposes to use user portrait as the basis of insider threat detection, it realizes the all-round construction of employee behavior information and improves the fine granularity of detection. The insider threat detection model can detect the threat employees in real time by introducing the sliding window mechanism to shift the portrait model of each employee adaptively. The detection model is validated in the CERT4.2 data set with good results.

Key words: insider threat detection; user portrait; adaptive

0 引言

内部威胁一直是网络安全领域中一个难以攻破的难题,对国家和企业都造成了很大的破坏和困扰。由 Ponemon 研究所和 IBM Security 联合制作的《2021 年数据泄露成本报告》[1]指出,相比 2020 年,2021 年的数据泄漏平均成本上升了 10%,并且泄漏的数据主要为客户的个人身份信息、客户数据、公司的知识产权以及员工的个人身份信息。报告指出恶意内部人员泄密造成的数据泄露事件的平均识别时间为 231 天,平均遏制时间为 75 天,在所有泄漏事件中排名第三,并且恶意内部人员所造成的经济损失占总损失的 8%。为了有效检测和识别企业内部的威胁员工,避免重要数据的泄漏和破坏,内部威胁检测系统的不断迭代和完善迫在眉睫。

用户画像是对用户全体特征的概括和描述,根据需要对用户的特定信息和行为进行抽象的一种标签化用户模型。用户画像的构成一般由静态属性和动态属性组成。静态属性用于记录用户不会经常发生变动的静态特征,动态属性一般多用于记录用户的各种行为数据以及经常变动的特征。虽然画像技术主要用于个性化服务和精准营销等商业领域,但越来越多的学者将此技术用于其他领域[2-5]。用

*基金项目:四川省科技计划项目(2020YFG0151);企业信息化与物联网测控技术四川省高校重点实验室项目(2021WZY01)

户画像对用户使用简化的标签描述用户所具有的特定特征,并实现对满足该特征的用户或用户群体的精准定位。具有特定特征需求在信息安全领域同样适用。国内外已有部分学者率先将用户画像技术运用于入侵检测技术和内部威胁领域当中,并取得了一定的研究成果。

在最近的调查中,文献[6]提出了一种基于多维特征的内部威胁检测混合模型,实验结果表明,ADAD和ATAD模型具有鲁棒性,混合模型明显优于两个分离模型。赵刚和姚兴仁[7]将用户画像技术用于入侵检测技术中,提出了基于用户画像的入侵检测模型,实现入侵检测粒度的细化。张建平[8]提出一种基于流量与日志的分析方法,构建用户关键行为的数据画像,实现了专用网络中用户关键行为监测的系统。郭渊博等[9]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,将局部描写与全局预测相结合,提高了检测准确率。钟雅等[10]将组织内的用户作为研究主体,将内部威胁检测与可视化相结合取得较好效果。虽然用户画像技术在入侵检测技术和内部威胁领域中取得一定的研究进展与效果,但大多数学者都只运用了用户画像技术中标签化模型的特点,没有将所有用户的行为信息整合进行检测,缺少对每一个员工行为画像的刻画,降低了检测的细粒度。为了提高内部威胁检测粒度的细化程度,本文将用户画像技术作为内部威胁检测的基础,并基于员工的动态行为信息构建内部威胁实时检测模型。

1 内部员工画像模型

企业数字系统中的用户行为在很大程度上取决于他们在组织中的不同角色或职位,及其负责的不同项目和任务。因此,用户在系统中的行为很可能反映他们需要定期执行的任务。将用户画像作为内部威胁检测的核心概念,根据员工的动态行为特征和静态属性构建出员工的用户画像模型进而提取出员工的潜在特征,画像的模型构建过程如图1所示。

1.1 员工静态画像

员工的静态画像一般用于刻画出员工的静态信息,如员工的邮件、职位、所在部门、团队、上司以及员工心理状态测评结果等。员工的静态画像虽不能直接用于内部威胁检测,但可用作辅助分析,便于对检测出为威胁员工的信息快速整理和概括,包括其在整个企业中的隶属关系等。

1.2 员工动态行为画像

内部威胁检测中最重要的信息就是员工行为数据信息,企业中对员工行为信息的收集一般都以格式化日志的形式进行,此类格式有助于大型重复性数据的记录和整理。用户画像模型构建之前需要对日志数据进行预处理并分片成以各员工为单位的数据结构。

1.2.1 员工行为画像的初始模型及构建规则

画像初始模型的构建需要将员工起初一段时间的行为数据都视作正常的行为操作。该模型的构建是对初始时间段的所有行为通过行为树及行为特征分类进行归纳和整合,行为树的结构如图2所示。图中的活动类型表示用户的行为类型,员工行为活动发生的时间段有上班前时间段、上班时间段以及下班后时间段,活动1~N表示在对应上班时间段行为发生的顺序以及行为特征类型。

模型中的行为特征类型用于对员工的行为信息进行分类,分类的标准是判断当前的行为信息是否为新的行为。通过判断此次行为产生的设备类型是否为未知设备、是否在该设备上发生过类似的行为以及该行为的属性是否一致来表示行为的特征

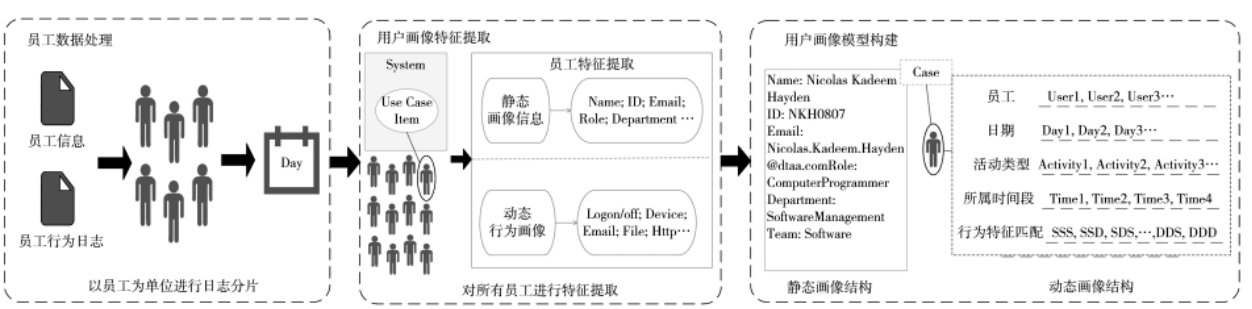

图1 内部员工画像模型构建

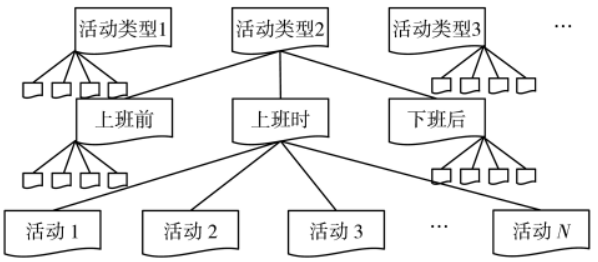

图 2 行为树结构

类型。具体特征表示结构如表 1 所示，其中员工以前未使用过的设备、未发生过的活动或者活动未产生过的属性标记为 N(New Event)，员工以前使用过的设备、发生过的活动以及活动产生过的属性则标记为 S(Same Event)。例如，员工 A 通过使用原来的设备 X 进行邮件发送行为，但是发送的对象是新对象，对此次行为的特征描述为 S-S-N。

表 1 员工单次行为画像的行为特征类型

序号	设备	活动	属性
1	S	S	S
2	S	S	N
3	S	N	N
4	S	N	S
5	N	S	N
6	N	S	S
7	N	N	S
8	N	N	N

员工的上下班是由员工所使用设备的开关机时间确定的，考虑到各员工上下班的情况不一，每个员工的上下班时间由各员工在初始时间段的上下班时间确定，上下班时间计算方法如下：

$$T_{on}=\frac{\sum_{am=1}^{n} t_{am}}{n}-30\ \text{min} \tag{1}$$

其中 T_{on} 表示员工上班时间阈值，t_{am} 表示该天设备的最早登录时间，n 表示初始时间段的天数。

$$T_{off}=\frac{\sum_{pm=1}^{n} t_{pm}}{n}+30\ \text{min} \tag{2}$$

其中 T_{off} 表示员工下班时间阈值，t_{pm} 表示该天设备的最晚注销时间，n 表示初始时间段的天数。

1.2.2 员工日行为画像模型构建

员工日行为画像模型主要用于记录员工一整天的活动轨迹同时作为内部威胁检测的基础模型和单位。模型的构建需对员工每次行为记录顺序进行特征提取，特征提取的过程也是对行为树的遍历过程。模型的结构如图 3 所示，主要构建流程如下：

(1)根据员工行为发生的时间顺序依次进行提取；

(2)对行为类型进行多分类；

(3)对行为所在时间段进行多分类；

(4)对行为进行特征提取；

(5)统计出当天各类行为在各个时间段产生的活动所对应特征类型数的数量。

1.2.3 员工行为画像模型的构建

员工行为画像模型是以日行为画像模型为单位构成的，是每个员工在初始时间段中生成的初始画像，该画像模型是对员工的行为模式和规律的初始刻画，同时其也是内部威胁检测中的初始模型。模型的结构如图 4 所示，其中 X 表示员工各特征类型出现的次数。模型的构建流程如下所示：

(1)确定员工初始时间段，此时间段中的员工数据需保证其行为记录均无异常；

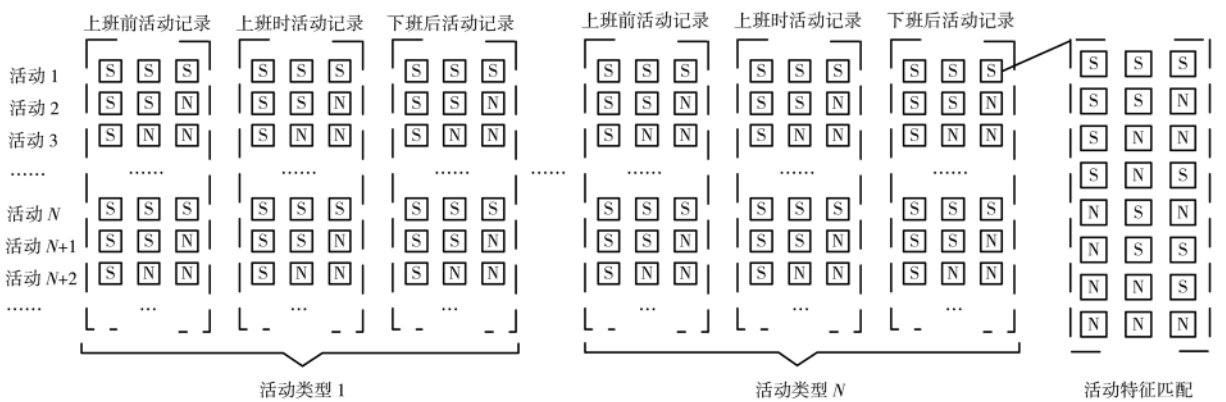

图 3 员工每日行为画像提取规则

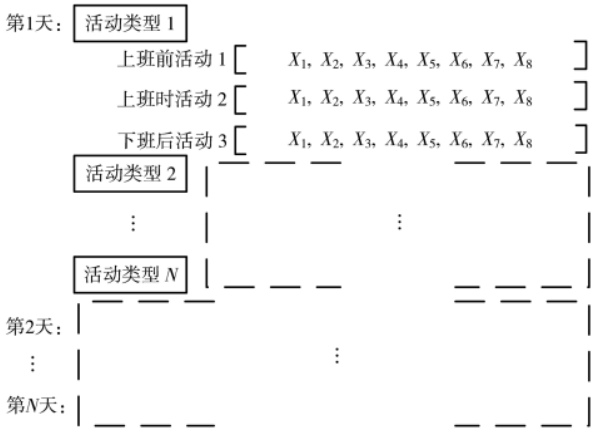

图 4 员工行为画像模型结构

(2)构建员工初始时间段内行为树;

(3)提取出员工初始时间段的每日行为画像记录;

(4)将员工在初始时间段的每日行为画像记录合并,将此作为员工的初始行为画像模型。其中每一行代表一天的行为特征的记录,而每一列则是以行为类型为单位的不同时间段的行为特征类型的数量的记录。

2 内部威胁实时检测模型

威胁检测模型的异常检测方法使用主成分分析(Principal Component Analysis,PCA)进行异常检测。为了保证正常画像模型的正确性并使模型随着用户的正常行为变化而变化,同时又能每天实时检测员工当天的异常情况,本文将员工画像的员工行为偏移模型采用滑动窗口机制。

2.1 基于用户画像的 PCA 检测模型

异常检测中通过对用户活动类型、活动时间段的正异常指标的设定来判断出员工是否具有异常。因画像模型产生的数据维度过高,实验使用 PCA 进行降维提取出行为特征中最主要的特征,同时通过计算行为前后数据点的偏差值大小检测当天行为是否异常,具体流程和规则如下。

将当前用户的行为画像模型用 X 表示:

$$X=[x_1,x_2,\cdots,x_n]^{\mathrm{T}},X\in R^n\times D,x_i\times R^{\mathrm{D}} \tag{3}$$

其中 x_i 表示员工日行为画像模型的向量形式,n 表示行数,D 表示特征数。

在进行 PCA 检测之前需对当前模型数据进行中心化和标准化,使得矩阵中的所有元素都能被缩放到 0~1 之间,消除不同量纲的影响。

中心化:

$$x_i^{'}=x_i-\frac{1}{n}\sum_i x_i \tag{4}$$

$$X_{\mathrm{ctz}}=[x_1^{'},x_2^{'},\cdots,x_n^{'}] \tag{5}$$

标准化:

$$x_i^{''}=\frac{x_i^{'}-\mu}{\sigma} \tag{6}$$

$$X_{\mathrm{std}}=[x_1^{''},x_2^{''},\cdots,x_n^{''}] \tag{7}$$

其中 μ、σ 分别表示 X 对应列上的均值和标准差。

将预处理之后的矩阵记为 X_{std},检测的主要步骤如下:

(1)计算协方差矩阵,并求出 C 的特征值$\{\lambda_1,\lambda_2,\cdots,\lambda_i\}$和特征向量$\{e_1,e_2,\cdots,e_i\}$,同时将特征向量按其特征值大小从上到下按列排列成矩阵,取前 k 列组成矩阵 P_k:

$$C=\frac{1}{n-1}X_{\mathrm{std}}^{\mathrm{T}}X_{\mathrm{std}} \tag{8}$$

(2)将高维数据 X_{std} 投影到低维数据 Y_k:

$$Y_k=X_{\mathrm{std}}P_k \tag{9}$$

其中,$Y_k=[y_1,y_2,\cdots,y_n]^{\mathrm{T}},Y\in R^n\times k,y_i\in R^D,k<<D$。

(3)求得不同维度中各数据点与其他数据的偏差:

$$d_{ij}=\frac{(y_i^{\mathrm{T}}\cdot e_j)^2}{\lambda_j} \tag{10}$$

其中 d_{ij} 表示该方向上的偏离程度,e_j 表示某特征向量,y_i 表示数据样本,λ_j 表示向量对应的特征值。

(4)根据各方向上的偏离程度求其综合异常得分

$$\mathrm{Score}(X_i)=\sum_{j=1}^{n}d_{ij} \tag{11}$$

(5)设置合适的阈值判断员工是否异常

$$\mathrm{Score}(X_i)<C \tag{12}$$

其中,C 为判断阈值。

2.2 员工行为画像的自适应偏移模型

员工初始画像模型构建完之后,画像模型的更新进入滑动阶段,实时检测员工异常原理如图 5 所示,其中 X 表示员工画像模型的特征。根据员工画像的构建规则,可将每个员工的画像模型转换成一个矩阵。

滑动窗口的构建需要对画像模型设置固定的天数作为滑动窗口的大小。画像模型中滑动窗口的规则如下:若异常检测模型检测出员工当天行为无异常,则会对该员工的画像模型进行偏移,保持员工的行为规律与画像模型同步;若检测出员工当天

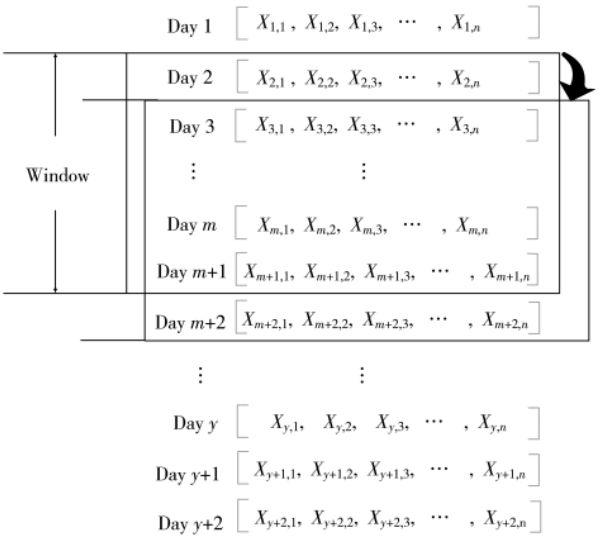

图 5 画像自适应偏移原理

有异常行为，为保证模型的正确性则不会对画像模型进行偏移。需要注意的是，画像模型的偏移并不会对员工的行为树产生影响。员工画像模型与异常检测模型息息相关，引入滑动窗口机制使得模型适应用户的行为习惯变化。

3 实验与结果分析

3.1 实验数据

为了验证检测模型的有效性，本实验使用公开数据集 CERT4.2 进行实验测试。CERT4.2 数据集记录了 1 000 名员工的资料信息以及在 500 天内所使用 PC 设备的所有活动，数据集结构和内容说明如表 2 所示，数据集的活动数目统计如表 3 所示。

表 2 数据集各个日志文件记录信息

文件名	文件记录说明
LDAP 文件夹	员工每个月的成员变动情况以及个人信息
logon.csv	员工登录和注销 PC 设备的型号和时间
device.csv	员工连接和断开存储设备的时间以及对文件的拷贝记录
file.csv	员工打开文件的时间、文件类型以及文件操作(复制、编辑、删除)
email.csv	员工接收/发送内/外部电子邮件的情况，包括发送时间、发件人、收件人、主题和关键内容信息
http.csv	员工浏览过的网站内容以及使用网站进行上传和下载的文件名和时间

表 3 CERT4.2 数据集的统计数据

员工数	恶意员工	活动	恶意活动
1 000	70	32 770 227	7 323

3.2 实验过程与结果分析

实验中，为了提高检测的细粒度，先将日志数据拆分成以员工为单位的数据结构，再对各个员工构建用户画像模型。初始画像模型构建完后，用数据模拟员工每天的行为活动，使用 PCA 检测模型进行实时检测。该数据集下内部威胁检测整体流程如图 6 所示。用户画像模型构建中每个员工每天的行为数据日志格式如表 4 所示。

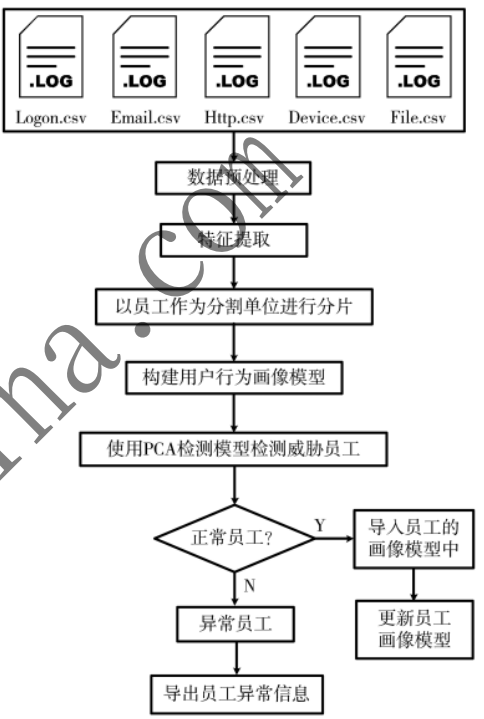

图 6 使用 CERT-4.2 数据集进行内部威胁检测流程

表 4 员工每日行为记录格式

活动类型	记录特征
logon	logon , user , pc , activity , date
file	file , user , pc , filename , activity , date
email	email , user , pc , from , to , size , attachment_count , date
http	http , user , pc , url , date
device	device , user , pc , activity , date

实验内部威胁检测模型测试中，将异常时间段个数的判断阈值设为 1、异常天数判断阈值设为 2。当把 PCA 降维数据所得平方和的标准化行为类型检测阈值设为 2.477 时，模型测试具有较好的检测结果。

基于用户画像的 PCA 内部威胁检测与非用户画像的 PCA 内部威胁检测及三种机器学习模型的

表5 各内部威胁检测方法结果对比

(%)

检测方法	威胁员工检测正确率	检测细粒度
基于用户画像的PCA方法	97.14	检测细粒度为威胁员工的异常时间段,检测正确率为82.75
基于非用户画像的PCA方法	78.57	检测细粒度为异常员工数
Random Forest	89.59	检测细粒度为异常员工数
Naive Bayes	91.96	检测细粒度为异常员工数
K-Nearest Neighbors	94.68	检测细粒度为异常员工数

实验对比结果如表5所示。实验检测到威胁员工的正确率为97.14%,检测到员工异常行为所在时间段的正确率为82.75%。实验结果证明了用户画像模型作为内部威胁检测基础的可行性。

4 结论

本文通过引入用户画像概念,将员工行为画像模型作为内部威胁检测的基础,根据员工的行为数据构建出员工行为树,并构建异常行为规则和检测阈值。同时,引入滑动窗口机制,使员工画像模型根据员工正常行为轨迹变化而产生偏移。最后,使用PCA进行特征降维和分类得出异常检测结果。模型以员工行为日期作为检测单位,增加了内部威胁检测的细粒度,能实时检测出发生异常的员工,实验结果证明模型具有较好的检测结果。后续研究将进一步提高检测的细粒度,引入更优的异常检测模型以提高检测的精度和效果。

参考文献

[1] WIDUP S,PINTO A,HYLENDER D,et al.2021 DBIR master's guide[EB/OL].(2021-XX-XX).https://www.verizon.com/business/resources/reports/dbir/2021/masters-guide/.

[2] YAO W,HOU Q,WANG J,et al.A personalized recommendation system based on user portrait[C]//Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science,2019:341-347.

[3] WANG H,TU Z,FU Y,et al.Time-aware user profiling from personal service ecosystem[J].Neural Computing and Applications,2021,33(8):3597-3619.

[4] CHEN T,WANG Y,YANG J,et al.Modeling multi-dimensional public opinion polarization process under the context of derived topics[J].International Journal of Environmental Research and Public Health,2021,18(2):472.

[5] BABU S.Detecting anomalies in users-an UEBA approach[C]//Proceedings of the International Conference on Industrial Engineering and Operations Management,2020:10-12.

[6] LV B,WANG D,WANG Y,et al.A hybrid model based on multi-dimensional features for insider threat detection[C]//International Conference on Wireless Algorithms,Systems,and Applications,2018:333-344.

[7] 赵刚,姚兴仁.基于用户画像的异常行为检测模型[J].信息网络安全,2017(7):18-24.

[8] 张建平,李洪敏,贾军,等.一种基于流量与日志的专网用户行为分析方法[J].信息安全研究,2020,6(9):783-790.

[9] 郭渊博,刘春辉,孔菁,等.内部威胁检测中用户行为模式画像方法研究[J].通信学报,2018,39(12):141-150.

[10] 钟雅,郭渊博,刘春辉,等.内部威胁检测中用户属性画像方法与应用[J].计算机科学,2020,47(3):292-297.

(收稿日期:2022-09-07)

作者简介:

李俊强(1996-),男,硕士研究生,主要研究方向:网络安全。

黄洪(1976-),男,博士,副教授,主要研究方向:系统安全测评、渗透测试、安全防护、大数据安全。

周子云(1995-),男,硕士研究生,主要研究方向:物联网安全。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部