计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190200379

内部威胁检测中用户属性画像方法与应用

钟　雅[1]　郭渊博[1]　刘春辉[2]　李　涛[1]
1 信息工程大学密码工程学院　郑州 450001
2 中国人民解放军 61213 部队　山西 临汾 041000
(17707404536@163.com)

摘　要　随着信息技术与互联网技术在企业组织中的广泛应用,企业安全面临着前所未有的挑战。大多数企业既面临着企业外部的攻击,也面临着内部人员的内部攻击。由于缺乏及时有效的检测手段,内部攻击对企业和组织造成的损害在一定程度上比外部攻击更加严重。在组织和企业内部,"人"是实施破坏行为的主体,是内部威胁检测中的主要研究对象。针对现有内部威胁检测中对内部员工完全隔离监管方法的相似威胁检测关联性低、检测效率低等问题,文中把研究重点从发现诱因转移到相似用户的聚类和监管上,以组织内的用户作为研究主体,提出了内部威胁检测中用户属性画像方法。该方法首先定义了画像相似度计算方法;然后,从用户性格、人格、过往经历、工作状态、遭遇的挫折等多方面着手,利用本体理论、标签式画像方法将多因素整合;最后,通过改进的 K-Means 算法实现用户聚类与分组管理,实现了潜在恶意用户共同监管的目的,减少了相似破坏多次发生的可能性。实验结果证明了所提方法的可行性,其为组织预防内部威胁提供了思路和方法。

关键词:企业安全;内部威胁;用户画像;群组管理;相似度计算;K-Means
中图法分类号　TP391

User Attributes Profiling Method and Application in Insider Threat Detection

ZHONG Ya[1],GUO Yuan-bo[1],LIU Chun-hui[2] and LI Tao[1]
1 Cryptography Engineering Institute,Information Engineering University,Zhengzhou 450001,China
2 Unit 61213 of The Chinese People's Liberation Army,Linfen,Shanxi 041000,China

Abstract　With the widely use of information technology and Internet technology in enterprise organizations,enterprise information security faces unprecedented challenges. Most companies are faced with both external and internal attacks. Due to the lack of timely and effective detection methods,the damage caused by internal attacks is more serious. As the conductor of malicious behaviors in organization and enterprise,human is the research object in insider threat detection. Aiming at the low correlation and low detection efficiency of the similar threat detection for the existing insider threat detection method,user attributes profiling method was proposed. In this paper,users in the organization were taken as the research subject,and the clustering and supervision of similar users were mainly studied. Firstly,the method of calculating the similarity of portraits is defined. Then,the ontology theory and tabular portrait method were used to integrate multiple factors,such as user personality,personality,past experience,working status,and setbacks. Similar users are clustered and managed in group by improved K-Means method,achieving the purpose of joint supervision on potential malicious ones,which reduces the possibility of similar damage occurring. Experimental results show that the proposed method is feasible and makes a way to combat the insider threat.

Keywords　Enterprise security,Insider threat,User profiling,Group management,Similarity calculation,K-Means

1　引言

内部威胁是指组织内部的合法员工、具有信息访问权限的合作方或第三方,违背了组织的安全策略,因恶意破坏或无意疏忽对组织或其内部资源的机密性、完整性和可用性造成损害的行为[1-4]。全球企业每年因内部威胁造成的损失占比越来越大,2015 年美国网络犯罪调查[5]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造成的损害。Verizon2018 数据违规调查报告[6]表明,28%的违规行为归咎于内部参与者。Dtex2018 内部威胁情报报告[7]指出,在过去的 1 年内,55%的组织遭受过多次内部攻击。由于缺乏及时有效的检测手

到稿日期:2019-02-28　返修日期:2019-04-17　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:国家自然科学基金(61501515)
This work was supported by the National Natural Science Foundation of China (61501515).
通信作者:郭渊博(yuanbo_g@hotmail.com)

段，内部威胁对企业和组织造成的损害一定程度上比外部攻击更加严重。

用户画像是一种从海量数据中抽取用户信息全貌，详细刻画用户内在需求和行为偏好的技术，在精准营销、推荐系统、搜索引擎等多种个性化定制场合得到了广泛应用。利用性格特征、人格特点、经济状况、感情状态、工作态度等属性特性对组织内员工进行画像，通过画像间相似度对员工进行聚类，可以实现对相似员工的分组管理。Legg 等提出了基于树形结构的员工、角色行为模型[8]，未提及员工属性特征的应用。Gamachchi 等提出了图处理单元与异常检测单元相结合的理论模型框架[9]，同样未利用员工属性特征。Nurse 等从员工内在特征出发，以攻击动机、员工技能树、工作态度、个人性格以及历史违规行为等多个方面对员工进行了描述，通过 3 个案例的解析，展示了内在特征在内部威胁检测中的作用[10]，但未提出内在特征建模方法。Liang 从人格、精神、心理等多项内在特征入手对恶意员工进行了调研[11]，研究了人格障碍、精神健康障碍、伦理问题、情绪特征、宗教信仰、负面经历以及经济状态等因素在员工实施内部攻击过程中发挥的作用，但未给出有效的度量标准。Liu 等[12]提出一种行为特征自动提取和局部全细节行为画像方法，该方法侧重个体行为，主要关注用户行为序列划分和全局业务状态转移，没有体现群体监管思想。

针对上述问题，本文将用户画像技术用于内部威胁检测，提出了用户属性画像方法。首先整理收集到的用户属性信息，概括出语义化、易处理的标签，利用本体[13]对标签进行表示、验证、推理和解释，形成用户属性画像。然后针对组织特点，定义了属性画像相似度计算方法，按照画像相似度对用户进行聚类，把研究重点从诱因发现转移到相似用户的聚类和监管上。当某用户执行异常操作后，加强对该用户所在群组其他用户的监管力度，为安全人员预防后续恶意行为提供了必要的数据支撑和决策支持。

本文第 2 节介绍本文方法的设计思路和各模块的实现方法进行；第 3 节介绍了在人工数据集上进行实验验证的基本流程；第 4 节对实验结果进行了分析，验证了该方法的可行性；最后对本文进行了总结和展望。

2　思路与方法

2.1　设计思路

在企业和组织内部，相同工作部门的员工因业务具有相关性，其业务操作有较强的相似性；而性格相似的员工的行为特征具有相似性。一般而言，某员工与恶意员工的相似性越大，执行恶意操作的可能性就越大。

实现真实有效的用户属性画像，需要细致调查组织内全部员工，与员工进行深入沟通，实时跟踪员工的日常表现。用户属性画像方法主要分为数据收集与整理、本体构建、相似用户聚类 3 个主要模块，如图 1 所示。

数据收集模块是基础，收集数据的真实程度、完善程度直接影响画像的准确性[14]。本体构建模块是属性画像的主要组成部分，是理清数据关系、提高数据表现力的关键步骤，本体构建的质量直接影响后续工作。相似用户聚类模块是体现属性画像价值的重要一环，通过定义相似度度量方法、改进聚类算法等实际手段，将具有相似属性画像的用户划分到同一群组，为实现相似用户共同监管打下了坚实的基础。

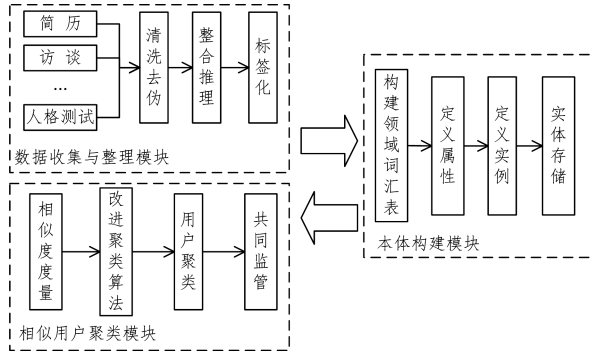


图 1　用户属性画像方法的设计思路

Fig. 1　Design ideas of user attribute profiling method

2.2　数据的获取与处理

用户属性信息的收集和确定是所有工作的基础，可从 3 个方面进行整理完善。1）简历信息，包括姓名、年龄、性别、感情状态、求职经历等相对静态不可变的信息；2）性格信息，如大五人格测试结果、心理评定结果、人际交往的深度和广度等；3）工作信息，如对待当前工作的态度、工作完成质量、对所属岗位的满意度、对上司和同事的满意度、个人技能与工作岗位的符合度、近期违反工作规章发生率以及是否因个人家庭、感情等原因影响工作进度等信息。其他属性信息可以通过调查问卷、心理测试、走访观察等多种方式完善。

收集完成后，须对数据进行清洗和预处理。首先清理无法真实表现用户特点的数据，理顺信息之间包含的关系，消除概念中可能存在的歧义。将无法完全量化的指标，按照实际情况划分等级，例如将个人能力分为胜任、刚好、不足 3 类，将态度分为积极、尚可、消极 3 类。随后根据数据间关系完成深度推理与整合，降低数据的信息冗余度。最后，归纳总结处理过的信息，形成用户属性标签。

2.3　用户本体构建方法与过程

本体是增强属性标签表现力的有效手段，是沟通属性信息与用户画像的直接桥梁，以员工为基本单位进行本体构建，主要包括以下几个关键步骤。

2.3.1　构建领域词汇表

领域词汇表标识并收集所有领域概念、属性和实例，这些词汇对应画像中的各类标签。领域词汇表一般包括类词汇表、属性词汇表等，其中属性词汇表又包括对象属性词汇表和数据属性词汇表。

建立领域词汇表有助于分析本体概念，去除冗余信息，保证涉及知识的完整性。在设计词汇表时，应结合系统需求考虑本体复用问题，例如“用户”类要继承“人”类中的基础属性，并根据需要增加新的属性。表 1 列出了用户属性画像领域内的部分词汇。

表 1 用户属性画像的类词汇表(部分)

Table 1 Class vocabulary for user attribute profiling (partial)

词名	类别	语义描述	所属类别
人	类	本体中所有人员的父类	Thing
组织	类	本体中所有组织信息的父类	Thing
业务	类	本体中所有业务活动的父类	Thing
员工	类	组织中的工作人员,人的子类	人
活动	类	员工完成业务过程中的一个操作,业务的子类	业务
部门	类	员工工作的部门,组织的子类	组织
职务	类	员工所担任的工作职务名称,组织的子类	组织

2.3.2 定义属性

属性包括对象属性(Objective Property)和数据属性(Data Property)两种。对象属性用来约束两个实例间的关系,定义域为类,值域为某个类的实例;数据属性用来约束类的实例,定义域为某个类的实例,值域为布尔型、字符串型、整型或者时间等[15]。表 2 列出了用户属性画像的部分属性。

表 2 用户属性画像的属性词汇表(部分)

Table 2 Attribute vocabulary for user attribute profiling (partial)

词条内容	类型	定义域	值域	语义描述	所属类别
任职	属性	员工	部门	—	对象属性
担任	属性	员工	职务	—	对象属性
执行	属性	员工	业务	—	对象属性
包含活动	属性	业务	活动	—	对象属性
上级部门	属性	部门	部门	—	对象属性
实施者	属性	活动	员工	—	对象属性
年龄	属性	人	整型	—	数据属性
感情状态	属性	人	字符串型	—	数据属性
工作态度	属性	员工	字符串型	—	数据属性
权限	属性	职务	字符串型	—	数据属性
时间戳	属性	活动	时间类型	—	数据属性

2.3.3 定义实例

如果类和属性是本体的“骨骼”,实例则是本体的“血肉”,类的实例通常包括实例名称、所属类别、实例说明等信息。本体中相关的实例示例如表 3 所列。将构建后的本体关系进行可视化,如图 2 所示。

表 3 本体中的实例(示例)

Table 3 Instances in the ontology (example)

实例名称	所属类别	说明
张三	员工	公司的一名员工
程序员	职务	工作职务之一
人事部	部门	企业组织架构中的一部分

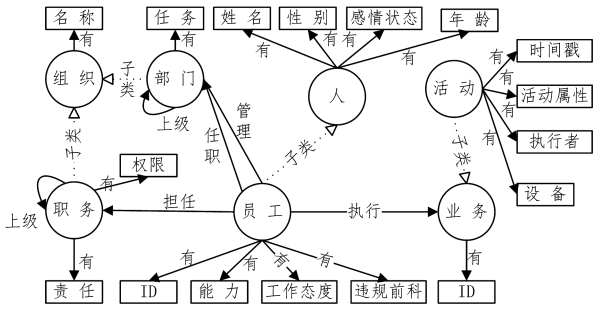


图 2 本体关系可视化

Fig. 2 Ontology visualization

2.4 画像相似性度量方法与相似用户聚类

计算用户画像间的相似度,根据画像相似度对员工进行聚类,可得到行为模式相近的员工群组。在基于本体的画像中,有些属性是定量的,有些属性是定性的,因此,计算画像间的相似度须结合定量相似度和定性相似度来进行。定量属性都有确定的数值,其相似度计算相对简单,可以采用传统的欧氏距离或余弦相似度进行度量。定性属性通过概念来表示,没有确定的数值,其相似度无法直接计算,文献[16]将概念名称、概念实例、概念属性相似度进行整合,提出了一种综合的本体相似度计算方法。Shi 等提出了一种基于局部密度、信息量和概念深度的混合算法[17]。

本文构建的本体具有结构相对简单、层次关系清晰的特点。因此,在现有本体相似度计算方法的基础上,采用相对成熟、便于解释的计算方法。根据属性定义的不同,相似度计算可从数据属性和关系属性两个方面讨论。数据属性可直接比较对应属性,而关系属性须结合该属性所在值域的层级关系和概念深度进行比较。

在计算员工 i 与员工 j 的相似度时,首先对员工类的所有数据属性进行比较,若 i 与 j 的对应属性的内容相同,则该相似度记为 1,否则记为 0。为提高相似度计算的容错率,可对某些属性相似度标准进行人工设定。例如,在大型成熟企业中,员工年龄层次明显,可设定员工年龄相差 3 岁以内为相同,在新兴创业型公司,员工年龄扁平化现象突出,则可设定员工年龄完全一致时为相同。

对于员工类中所有的对象属性,需要根据该属性所在值域的层次关系进行计算。例如,员工类的任职属性的值域为组织类的子类部门。部门间存在上下级关系,其树状组织架构关系如图 3 所示,同一层次部门的任务具有相似性,两部门在树状架构中重合度越高,其相似度越大。将完全重合记为 1,完全不同记为 0,其余情况下,根据重合度在全局中的比例确定。例如“研发管理—工程部—电子工程组”与“商业研发—工程部—电子工程组”的相似度为 2/3,“研发管理—工程部—电子工程组”与“商业研发—工程部—测试评估组”的相似度为 1/3。员工类的担任属性的值域为组织类的子类职务,职务类与部门类具有类似的隶属关系和树状架构,其相似度计算方式与部门类相同。

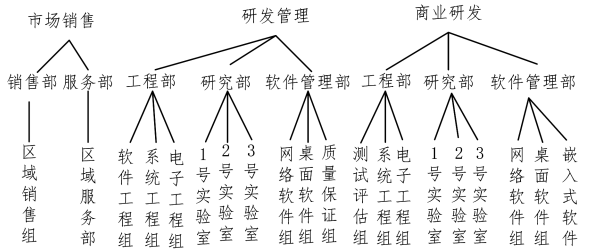


图 3 部门隶属关系图(部分)

Fig. 3 Department affiliation diagram (partial)

设员工 i 与员工 j 参与比较的属性总数为 n,其综合相似度如式(1)所示:

$$sim(U_i,U_j)=\sum_n \omega_n S_n^{i,j} \tag{1}$$

其中,$S_n^{i,j}$ 指两员工每个属性的相似度;ω 指属性在相似性计算中所占权重,$\sum_n \omega_n=1$,不同组织可根据实际情况对 ω 进行调整。

基于相似度计算方法，采用改进的 K-Means 聚类算法对用户属性画像进行聚类。基于相似用户行为模式相近的假设，对相似用户进行聚类，可以得到行为模式相近的用户群。

3　实验验证

3.1　实验数据

由于企业数据的机密性、隐私性等，目前无法获取真实企业中的数据进行训练和测试。因此，本文使用认可度较高的 CMU-CERT 集成数据集作为实验验证数据源[18]。CMU-CERT 数据集是由美国国防部高级研究计划局(Defense Advanced Research Projects Agency，DARPA)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，该数据集还包含了大量正常的背景数据。

CERT 数据集包含企业内部 4 000 名用户 500 天的所有活动记录，部分记录为攻击活动。在用户属性信息方面，该数据集提供了用户角色、担任职务、工作部门等工作信息，并提供了大五人格评测得分的信息。具体内容如表 4 所列，日志格式文件如表 5 所列。

表 4　CERT 数据集用户属性信息

Table 4　Information of user attribute in CERT data set

类别	名称	说明	备注
工作信息	角色	用户的工作头衔	按照部门的体量大小自上而下降序排列
工作信息	项目	用户当前负责的项目	
工作信息	业务单元		
工作信息	功能单元		
工作信息	部门		
工作信息	小组		
工作信息	监管人		
人格信息	开放性 O	想象、情感、智能、求异等特质	
人格信息	责任性 C	胜任、公正、条理、自律等特质	OECAN 得分总和为 100，某项得分越高，该项特质越突出
人格信息	外倾性 E	热情、果断、乐观、冒险等特质	
人格信息	宜人性 A	信任、直率、利他、谦虚等特质	
人格信息	情绪性 N	焦虑、压抑、敌对、冲动等特质	

表 5　日志文件格式(以 psychometric.csv 部分数据为例)

Table 5　Format of log file(taking part of the data in psychometric.csv as examples)

employee_name	user_id	O	C	E	A	N
Nicholas Fletcher Pruitt	NFP2441	34	39	38	36	21
Abraham Dante Rodgers	ADR1517	36	39	13	19	27
Medge Wilma Blackburn	MWB4000	27	14	44	22	34
Meghan Laurel Salazar	MLS2856	35	49	22	45	28

大五人格模型是由 *Ernest Tupes and Raymond Christal* 于 1961 年提出，经过众多独立学者从人类行为学描述符的因子分析等多个不同侧面补充、完善、论证后得到的目前学术界公认的相对合理的人格描述模型。大五人格利用开放性、责任性、外倾性、宜人性、情绪性 5 种特质描述人格的绝大多数方面，依托个性心理和心理测量科学，客观分析人们在行为、动机、态度和期望方面的个性差异和特长，尤其在工作场合的人的表现和合作方面效果良好。

3.2　实验过程

为验证本文所提方法的可行性，我们利用 Python 语言开发了原型测试程序。测试环境为 CentOS 7，CPU 为 Intel i7-4790 @3.60 GHz，RAM 为 16 GB，硬盘为 1 TB 机械硬盘。

实验前，先对数据集中的日志数据进行预处理，去除本文不需要的内容，并对缺失内容进行填充，对错误信息进行清理，形成表 5 所示的实验输入数据。

利用输入数据对本体进行构建。构建过程中，以员工本体(Users)为主，以部门本体(Department)、职务本体(Role)为辅，用户静态属性以字符串形式表示，人格数据以五元整数数组的形式表示。为方便后续相似度计算，本文对部门本体、职务本体的上下级关系进行了分析，发现不同的功能单元中包含部分职能相似的部门和小组，故假定任职于不同功能单元中的职能相似部门或小组的用户具有相似性。

属性间相似度计算需考虑属性表示类型。在大五人格相似度计算过程中，相同特质间得分越相近，其相似度越接近。设用户 i 和 j 的人格属性得分分别为 $psy_i=[o_i,c_i,e_i,a_i,n_i]$ 和 $psy_j=[o_j,c_j,e_j,a_j,n_j]$，则相似度计算式为：

$$Sim_{psy}(i,j)=1-\sum_{k=1}^{5}\omega_k\,|psy_i(k)-psy_j(k)|/200 \tag{2}$$

其中，$psy_i(k)$ 表示属性得分数组中第 m 个数值，ω_k 表示第 k 个数值在整个人格属性相似度计算过程中的权重。企业可根据组织特点，突出某项人格特质在比较过程中的重要性。为方便计算，此处认为 5 项特质得分在相似度计算过程中所占比重相同。

字符串类型属性相似度采用 Levenshtein Ratio 进行计算。Levenshtein Ratio 是计算文本相似度常用的指标之一，其计算式为：

$$r=\frac{(sum-ldist)}{sum} \tag{3}$$

其中，sum 是指 $str1$ 和 $str2$ 字串的长度总和，$ldist$ 是指类编辑距离。编辑距离指两个字串之间，由一个转成另一个所需的最少编辑操作次数[19]。编辑操作包括字符替换、插入、删除。一般来说，编辑距离越小，两个串的相似度越大。而在类编辑距离计算中只有删除、插入操作，替换操作分解为删除加插入两步操作。

相似用户聚类采用基于相似度式(1)改进的 K-Means 聚类算法。该数据集包含 4 000 名员工，涉及 9 个功能单元、46 种职务、多种不同的人格特质。为使聚类群组更具有代表性，需要确定聚类分组数 K。本文对评价聚类效果的均方差(Computer emergency response teams，MSE)指标进行了改进，得到聚类平均相似度指标(Mean Similarity，MS)。

当聚类数为 K 时，聚类完成后，K 个聚类中心分别为 $U_{i,0}(i=1,2,\cdots,K)$。计算每个聚类中各用户 $U_{i,j}$ 与其聚类中心点 $U_{i,0}$ 的相似度，所有用户的相似度的平均值即为该次聚类的平均相似度，如式(4)所示：

$$MS=\sum_{i=1,j=0}^{K,J_i} sim(U_{i,j},U_{i,0})/\sum_{i=1}^{K}J_i \tag{4}$$

其中，J_i 为第 i 个群组所包含用户的总数。

从式(4)可以看出，当用户数量一定时，K 越大 MS 值越大；当 K 等于用户总数量时，$MS=1$。然而，K 值越大计算越复杂，且当 K 值增大到一定程度时，相似群组中的用户个数急剧减少。当群组中平均用户数小于某数值阈值时，便失去

了相似用户聚类的意义。根据 K-MS 值的变化规律，找到 *MS* 增幅变化最小的 *K* 值，即为合理 *K* 值。

随着时间的推移，用户的人格属性不会发生明显的变化，但其担任职务、所在部门会发生变化，须定时对用户属性进行更新，并重新确定更新后用户的相似群。数据集以月为单位对用户的职务、部门等信息进行了统计，提供了 18 个 LDAP 文件，因此本文设定以月为单位对用户属性进行更新。最后，利用数据集提供的恶意用户标签，对所提方法可行性进行验证。

4 结果分析

选择合适的 *K* 值是进行 K-Means 聚类的关键，是保证用户聚类准确性的前提。实验过程中，*K* 从 5 开始以 1 为步长顺序递增。每个 *K* 值下，独立进行 10 次聚类操作，取其中的最大 *MS* 值绘出 K-MS 折线图。

从图 4 中可以看出，随着聚类数 *K* 值增大，*MS* 值不断增大，当 *K* 值增大到一定程度后，*MS* 的增幅逐渐减少。根据手肘法 *K* 值确定原则，选择 *MS* 增幅最大的 $K=20$ 作为后续实验的最佳聚类数。随后，对所有用户根据其画像相似度进行聚类，相似用户被划分至同一个分组。每个分组包含的用户数如表 6 所列。图 5(a)展示了各分组包含的用户数量，图 5(b)展示了各分组人数在总人数中所占的比例。从图 5 可以看出，聚类完成后，每个群组内的用户数趋于平均，其中分组 16 人数最少，为 39 人，占总人数的 0.975%，分组 18 人数最多，为 387 人，占总人数的 9.675%。

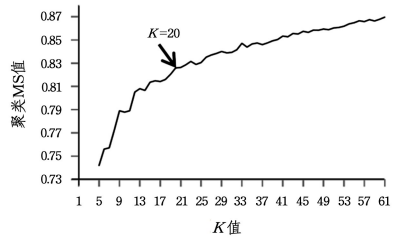


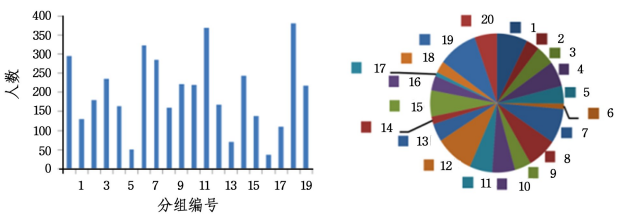
图 4 K-MS 折线图

Fig. 4 Line chart of K-MS

表 6 所有用户分组情况

Table 6 Grouping situation of users

编号	分组				…		
	0	1	2	3	…	18	19
人数	296	118	180	236	…	387	250
比例/%	7.40	2.95	4.50	5.90	…	9.68	6.25



(a)各分组包含的用户数 (b)各分组人数在总人数中所占的比例

图 5 所有用户分组情况

Fig. 5 Grouping situation of all users

为证明用户属性画像在内部威胁检测中的作用，根据 CERT 数据集提供的攻击者标签，将攻击者所在分组情况进行统计。数据集中包括攻击者 28 人，占总用户数的 0.7%，攻击者所在分组及其所占比例如图 6 所示。

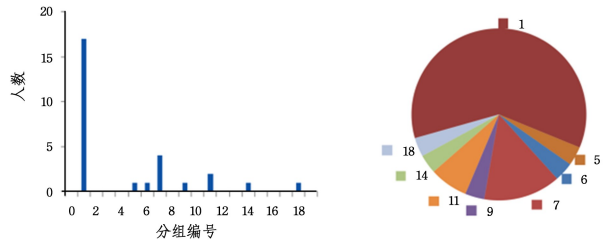


图 6 攻击者分组情况

Fig. 6 Grouping situation of attackers

图 6 中，分组 1 包含攻击者 17 人，占攻击者总数的 60.7%，占该组总用户数的 14.4%；而分组 18 包含攻击者 1 人，占该组总用户数的 0.26%。编号 0,2,4 等 12 个分组不包含攻击者。从图 6 可知，大多数攻击者相似度较高，利用用户属性画像相似度聚类，可以发现大量相似的恶意用户。

将 28 名攻击者进行的全部破坏行为按发生时间的先后顺序进行排列，结合攻击者所在用户组画出破坏行为发生时间图，如图 7 所示。

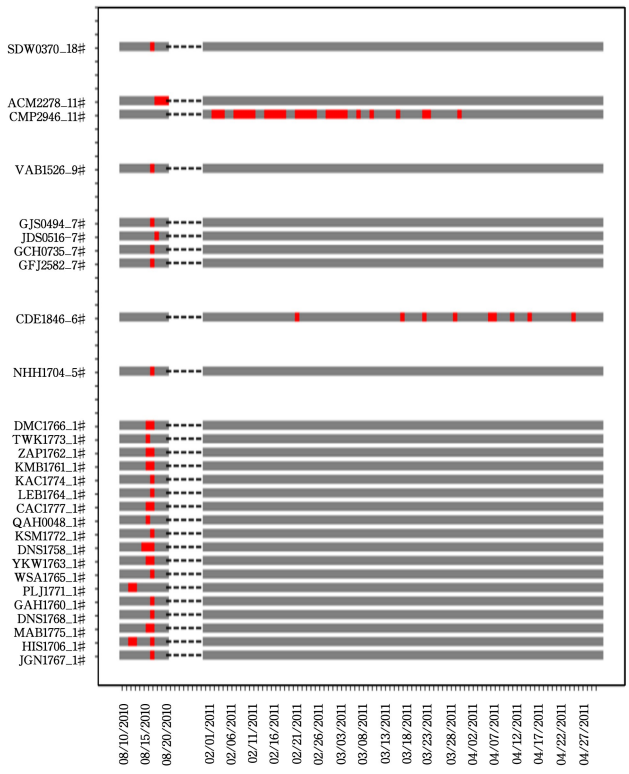


图 7 恶意用户破坏行为发生时间图

Fig. 7 Occurrence time of destructive behavior for malicious users

本文提出了基于用户属性画像相似度的用户聚类和共同监管方法，而内部威胁检测领域并未对该方法进行研究，且本文尚未提出定量刻画方法有效性的指标，无法与业界结果形成对比。但从图 7 可以看出，若用户 PLJ1771 和 HIS1706 发动恶意行为，则加大对 1 号分组用户的监管力度，可降低后续恶意行为发生的概率。在分组 7、分组 11 中，该方法同样有效。因此，利用属性画像相似度对相似用户进行分组管理，当

发现攻击时对该用户所在群组加强监管，可有效降低该组用户继续攻击的可能性，进而降低后续恶意行为对组织造成的损失。

结束语　内部威胁攻击隐蔽性强、破坏性大，直接威胁到企业的核心利益，因其复杂性和企业数据的隐私性问题，检测效果一直不佳。本文将组织内的用户作为研究主体，把用户画像技术移植到内部威胁检测过程中，从用户性格、人格、过往经历、工作状态、遭遇的挫折等多方面着手，通过本体理论、标签式画像方法将多因素进行整合，利用画像相似度对用户进行聚类，突出与恶意员工行为动机高度相似的“高危”用户。在内部威胁检测实践中，加强对“高危”用户的监管力度，可降低内部威胁的发生率，减少内部攻击对企业造成的损失。

参 考 文 献

[1] BISHOP M, GATES C. Defining the insider threat[C]// Proceedings of the Cyber Security & Information Intelligence Research Workshop. 2008.

[2] PATZAKIS J. New incident response best practices: Patch and proceed is no longer acceptable incident response [J]. Guidance Software, Pasadena, CA, Tech. Rep, 2003(9): 97-105.

[3] WARKENTIN M, WILLISON R, JOHNSTON A C. The Role of Perceptions of Organizational Injustice and Techniques of Neutralization in Forming Computer Abuse Intentions[C]// AMCIS 2011. Detroit, Michigan, USA: DBLP, 2011.

[4] PREDD J, PFLEEGER S L, HUNKER J, et al. Insiders behaving badly [J]. IEEE Security & Privacy, 2008, 6(4): 66-70.

[5] CSO Magazine, U. S. Secret Service, CERT Division of the Software Engineering Institute, et al. 2015 U. S. state of cybercrime survey [OL]. https://www.pwc.com/us/en/increasing-it-effectiveness/publications/assets/2015-us-cybercrime-survey.pdf.

[6] Verizon. 2018 Data Breach Investigations Report [OL]. https://www.verizonenterprise.com/resources/reports/rp_DBIR_2018_Report_execsummary_en_xg.pdf. 2018.

[7] Dtex Systems. 2018 insider threat intelligence report[OL]. https://www.dtexsystems.com/2018-insider-threat-intelligence-report. 2018.

[8] 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, 2017, 11(2): 503-512.

[9] GAMACHCHI A, SUN L, BOZTAS S. A Graph Based Framework for Malicious Insider Threat Detection[J]. arXiv: 1089.00141, 2017.

[10] NURSE J R C, BUCKLEY O, LEGG P A, et al. Understanding insider threat: A framework for characterising attacks[C]// IEEE Security and Privacy Workshops. ACM, 2014: 214-228.

[11] LIANG N. Characteristics of Malicious Insiders and Their Relationships with Different Types of Malicious Attacks[D]. Stillwater: Oklahoma State University, 2017.

[12] GUO Y B, LIU C H, KONG J, et al. Research on User Behavior Patterns Profiling in InsiderThreat Detection [J]. Journal of China Institute of Communications, 2018, 39(12): 145-154.

[13] ABBESH, BOUKETTAYA S, GARGOURI F. Learning ontology from Big Data through MongoDB database[C]// Computer Systems & Applications. IEEE, 2016.

[14] QIU R C, ANTONIK P. The Mathematical Foundations of Data Collection[M]// Smart Grid using Big Data Analytics: A Random Matrix Theory Approach. 2017.

[15] JIA W Y. Research on personalized recommendation algorithm of agriculture information based on group users' portrait[D]. Xianyang: Northwest A&F University. 2017.

[16] ZHANG Z P, TIAN S X, LIU H Q. Compositive Approach for Ontology Similarity Computation[J]. Computer Science, 2008, 35(12): 142-145.

[17] SHI B, FANG L, YAN J, et al. Ontology-Based Measure of Semantic Similarity between Concepts[C]// IEEE Computer Society. Xiamen, 2009: 109-112.

[18] US-CERT. Insider Threat Tools [EB/OL]. http://www.cert.org/insider-threat/tools/index.cfm, 2014-10-20.

[19] LUO Y G, LI X, JIANG T H, et al. Uyghur Lexicon Normalization Method Based on Word Vector[J]. Computer Engineering, 2018(2): 220-225.

ZHONG Ya, born in 1995, postgraduate. Her main research interests include insider threat detection and anomaly detection.

GUO Yuan-bo, born in 1975, Ph.D, professor, is member of China Computer Federation. His main research interests include network attack and defense confrontation.