计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2017,26(12):220−226 [doi: 10.15888/j.cnki.csa.006064]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于相似度分析的电力信息内网用户行为异常预警方法①

金倩倩[1], 陈春霖[2], 于晓文[1], 廖　鹏[1]

[1](南瑞集团公司 (国网电力科学研究院), 南京 210003)
[2](国家电网公司, 北京 100031)

摘　要: 用户作为网络的重要主体, 对其进行行为分析是掌握网络安全状态的重要手段, 且在异常检测中对于潜在威胁挖掘和预警具有重要的意义. 本文从电力信息内网同类型用户间行为存在相似性的角度考虑, 基于时间行为序列建模对单个用户的行为进行描述, 并通过用户行为相似情况的自学习建立用户间的关联, 以行为相似偏差实现异常分析, 同时考虑用户基础属性的变化实现异常预警判定. 通过模拟实验, 该方法能够有效地利用行为序列间的相似度发现潜在的异常行为并进行预警.

关键词: 用户行为分析; 行为相似度; 时间序列; 安全预警

引用格式: 金倩倩,陈春霖,于晓文,廖鹏.基于相似度分析的电力信息内网用户行为异常预警方法.计算机系统应用,2017,26(12):220–226. http://www.c-s-a.org.cn/1003-3254/6064.html

Early Warning Method of Anomaly User Behavior Based on Similarity Analysis in Power Intranet

JIN Qian-Qian[1], CHEN Chun-Lin[2], YU Xiao-Wen[1], LIAO Peng[1]

[1](NARI Group Corporation State (Grid Electric Power Research Institute), Nanjing 210003, China)
[2](State Grid Corporation of China, Beijing 100031, China)

Abstract: As an important subject of the network, the behavior analysis of users is an important means to grasp the network security state and has a significant meaning on potential threat mining and early warning. Considering that users of similar roles in the power intranet have similar behaviors, this paper describes the behavior of individual users based on time sequence and builds behavior relevance among the users by self-learning of the similarity of users' behaviors to achieve abnormality analysis by means of behavior similarity deviation. Meanwhile, changes of users' basic attributes are considered to achieve abnormality early warning judgment. Simulation experiments show that the method can discover abnormal behavior and perform early warning by effectively using the similarity analysis of the behavioral sequences.

Key words: user behavior analyze; behavior similarity; time sequence; security early warning

大量针对高级威胁的外部攻击的分析研究表明, 外部的攻击要真正达到目的必须经过"内网行走"才能接触到敏感数据等; 大量的安全事件是内部的恶意或无意员工造成, 或长期的潜伏或离职意向前的突发行为, 或内外结合造成. 可见, 某种程度内部攻击更难防范, 所以在加强外部防御的同时应更加重视对内网的异常监测和安全防护. 异常行为发现是进行内网安全监测的重要触发点, 用户作为内网行为主体, 对其行为分析是内网中异常行为发现的重要手段. 用户行为分析是指从网站或者网络端口获得相关的网络流数据, 采用统计分析等方法对数据进行处理, 研究网络用户的特点、构成及其行为活动上所表现出来的规律, 借以

① 基金项目: 国家电网公司科技项目 (SGFJXT00YJJS1600064)
收稿时间: 2017-02-27; 修改时间: 2017-03-16; 采用时间: 2017-03-23

预测并控制可能的违规行为[1]. 掌握用户的行为习惯, 对于预测用户上网行为及异常行为发现具有重要的意义.

国家电网公司为了满足特大型电力企业信息化建设的安全保障需求, 结合智能电网业务系统的建设需要, 基于国家等级保护各项基本要求, 在传统电力二次系统安全防护实现生产控制大区和管理信息大区单向隔离的基础上, 通过技术改造与创新将国家电网公司管理信息网划分为信息内网和信息外网并实施有效的隔离[2,3]. 信息内网针对电力业务、办公、运维等提供网络服务, 并具备明确的业务功能分工. 基于此网络环境, 内网用户的网络访问行为存在如下特征: 1) 不同部门、不同岗位员工的网络访问行为会因为角色分工体现明显的差异性和相似性. 例如, 电网运行维护人员会频繁访问运维监控系统, 但不会访问人财物系统. 2) 由于工作流程的标准化, 使员工对于不同系统或设备的访问顺序存在一定的规律. 例如, 电网运行维护人员每日会查看邮件领取待办任务, 通过监测系统关注运行状态, 对问题进行工作票下发, 且通过堡垒机等对设备进行状态确认或检修. 所以, 若出现超出角色职责、违背日常行为习惯或访问规律的行为, 同时不存在该员工岗位变动的, 即可认为用户行为异常, 网络中的系统存在被攻击的风险.

目前, 针对内部网络中行为的异常检测方法多针对基于网络流行为分析的方法[4-9]和基于网络协议异常的分析方法[10,11], 通过统计、向量机等手段对流量和协议行为进行建模从而检测异常. 基于用户行为的分析多针对主机行为进行研究, 如通过主机审计命令进行网络行为建模[12-14]、通过正常行为聚类方法判定异常[15,16]、通过网络访问流量分析进行用户行为建模从而检测异常[17,18]等. 但是, 以上分析方法只考虑了用户单次行为模式, 忽略了用户多次访问或操作间的时序关联关系, 且未考虑多个用户间的关系以及用户岗位等属性变更带来的影响, 在电力信息内网环境下不能完整地对用户行为模式进行描述.

针对上述问题, 本文提出一种基于相似度分析的用户行为异常预警方法, 基于网络流量数据按照时间提取用户访问行为序列, 并引入不同用户间的行为序列相似度和相关系数的概念, 通过用户行为序列相似度描述用户间的关联, 同时考虑用户属性的变化, 对异常行为进行发现和安全预警.

1 用户行为序列模式描述

1.1 基于时间的行为序列

假设有两个用户分别访问目标主机 A、B、C、D, 在一定时间段内, 用户 1 先后访问目标主机 A、B、D、C, 用户 2 先后访问目标主机 B、A、D, 两个用户的访问序列示意如图 1 所示.

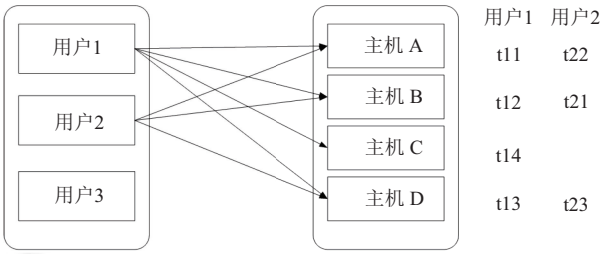


图 1 用户访问行为序列示意图

图 1 中, t_{11} 表示用户 1 发生第 1 次访问行为的时间, 即用户 1 访问主机 A 的时间, t_{12} 表示用户 1 发生第 2 次访问行为的时间, t_{21} 表示用户 2 第 1 次访问行为的时间, 以此类推, 形成该分析场景下完整的用户访问行为的时间序列. 基于用户访问行为时间序列, 将用户 i 行为序列表示为 $ST_i=(Tt_{i1}, Tt_{i2}, Tt_{i3}, \ldots, Tt_{in})$, 其中, n 表示用户根据时间先后发生访问行为的次序, Tt_{in} 表示用户在 t_{in} 时间发生的具体的访问行为. 本文采用基于时间的用户行为访问序列描述单个用户的用户行为, 作为异常检测和预警的输入.

1.2 最大公共行为子序列

(1) 子序列

若给定序列 $X=(x_1, x_2, x_3, \ldots, x_n)$, 则序列 $Z=(z_1, z_2, z_3, \ldots, z_n)$ 为 X 的子序列的规则为: 存在一个严格递增的下标序列 $(i_1, i_2, i_3, \ldots, i_k)$, 使对于所有的 j=1, 2, …, k 有 $z_j=x_i$, 设起始下标为 1.

(2) 最大公共子序列

给定两个序列 X 和 Y, 当序列 Z 既是 X 的子序列又是 Y 的子序列, 则 Z 是序列 X 和 Y 的公共子序列. 其中 Z 最长的序列称为 X 和 Y 的最大公共子序列 (Longest common subsequence, LCS).

最大公共子序列的最优子结构特性: 设 $X=(x_1, x_2, x_3, \ldots, x_m)$ 和 $Y=(y_1, y_2, y_3, \ldots, y_n)$ 为两个序列, $Z=(z_1, z_2, z_3, \ldots, z_k)$ 是它们最大公共子序列, 则应满足如下特性.

1) 若 $x_m=y_n$, 则 $z_k=x_m=y_n$, 且 z_{k-1} 是 x_{m-1} 和 y_{n-1} 的

最大公共子序列;

2) 若 $x_m \neq y_n$ 且 $z_k \neq x_m$, 则 Z 是 X_{m-1} 和 Y 的最大公共子序列;

3) 若 $x_m \neq y_n$ 且 $z_k \neq y_n$, 则 Z 是 X 和 Y_{n-1} 的最大公共子序列. 由最优子结构的特性, 结合用户行为序列定义, 可以得到求解用户行为序列最大公共子序列的过程, 如图 2 所示.

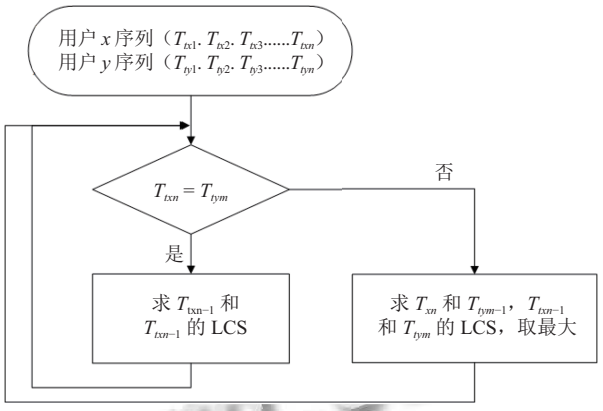


图 2 用户行为序列最大公共子序列计算流程

用 c[i][j]表示用户 x 和用户 y 的最大行为公共子序列, $ST_x=(Tt_{x1}, Tt_{x2}, Tt_{x3}, \ldots, Tt_{xn})$ 和 $ST_y=(Tt_{y1}, Tt_{y2}, Tt_{y3}, \ldots, Tt_{ym})$, 则有下列公式:

$$c[i][j]=\begin{cases} 0 & i=0, j=0 \\ c[i-1] & i,j>0, Tt_{xi}=Tt_{yj}, i<n, j<m \\ \max(c[i][j-1], c[i-1][j]) & i,j>0, Tt_{xi}\neq Tt_{yj}, i<n, j<m \end{cases}$$

由此求得两个用户之间的最大公共行为子序列.

1.3 行为序列相似度矩阵

根据用户行为最大公共子序列, 可计算出不同用户之间的行为序列相似度, 表示不同用户间的行为相似性.

设用户 x 行为序列 $ST_x=(Tt_{x1}, Tt_{x2}, Tt_{x3}, \ldots, Tt_{xn})$, 序列的长度为 n, 用户 y 行为序列 $ST_y=(Tt_{y1}, Tt_{y2}, Tt_{y3}, \ldots, Tt_{ym})$, 序列长度为 m, 求出两者最大公共子序列 $C=(c_1, c_2, c_3, \ldots, c_h)$, 序列长度为 h, 则用户 x 与用户 y 行为序列相似度的计算公式为:

$$Similarity(ST_x, ST_y)=\frac{h\times 2}{m+n}$$

对于 n 个用户, 通过两两相似度计算, 可得 n×n 的上三角矩阵, 称为用户行为序列相似度矩阵 $M(ST)_{n\times n}=(m_{xy})_{n\times n}$, 计算公式为:

$$m_{xy}=\begin{cases} Similarity(ST_x, ST_y), & x\le y \\ 0, & x>y \end{cases}$$

根据用户行为序列相似度矩阵, 可很清晰地看出每两个用户之间在一定时间段内的访问行为相似程度.

本文采用基于 Jaccard 算法[19]改进的相似度计算方法 Common_Jaccard 算法进行用户行为相似度计算, 能够更准确地描述用户行为相似性. 假定用户序列 A 及用户序列 B, len() 为求序列的长度, 最大公用子序列为 C, 则使用 Common_Jaccard 算法计算用户 A 和用户 B 的相似度 α_{cj} 的公式为:

$$\alpha_{cj}=\frac{len(C)}{len(A\cup B)}$$

1.4 用户行为相关系数

通过分析一段时间内 (前 n 个时间窗) 行为序列相似度的变化, 可以得到该时间段内, 访问行为最相近的用户组合或用户类. 平均相似度 α_{avg} 越大, 相似度变化越小, 则这两个用户关系越相近. 假设相似度方差为 α_{dx}, 则两个用户的行为相关系数为:

$$RC=\frac{\alpha_{avg}}{\alpha_{dx}}$$

可见, 两个用户之间相关系数 RC 越大, 则这两个用户的行为关系越相近. 有了相似度 α 和相关系数 RC, 就能够更精确的描述用户之间行为相似程度, 反应用户之间的关系, 从而实现异常行为分析.

例如, 在完成前 n 个时间窗行为序列相似度训练后, 可得两个用户间的相关系数平均值 RC_{avg} 和相关系数方差 RC_{dx}, 以 $RC_{avg}\pm RC_{dx}$ 作为后续检测的正常结果参考上下限, 若用户间相关系数超出参考上下限, 则判定出现异常的用户行为.

1.5 用户基本属性

本文采用六元组描述电力内网用户基本属性, User={name, ip, department, post, role, latestupdatetime}, 其中, name 表示姓名, ip 表示用户的绑定终端的 ip 地址, department 表示用户当前所在部门, post 表示用户当前职位, role 表示用户的角色分工, latestupdatetime 表示基本属性最近更新时间.

用户基本属性是对通过行为相似度分析发现的异常进行关联判断最终生成预警的关键要素.

2 基于行为序列的异常分析

2.1 系统架构

基于行为序列模式构建, 本文提出了一种基于相似度分析的电力信息内网用户的行为异常预警方法. 通过数据预处理、行为序列模式挖掘、行为异常分析实现异常行为发现, 并通过关联用户属性判断预警的生成. 系统架构如图 3 所示.

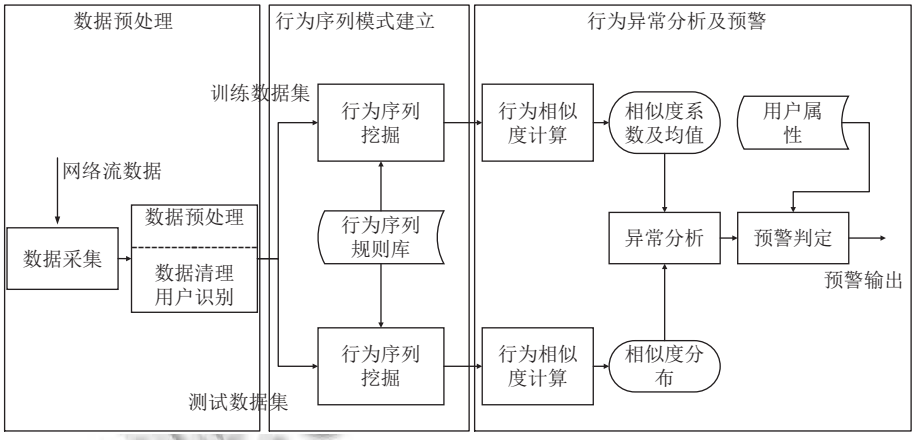


图 3 基于相似度分析的行为异常预警系统结构

下面分别从这三个阶段对本文提出的内网用户异常行为检测方法进行详细描述.

2.2 数据预处理

原始数据来源于所监测网络中的网络流报文, 数据预处理的目的是为了减少所捕获网络流数据中的无效数据, 包括剔除原始数据中的冗余信息、错误信息及与分析不相关的用户行为数据, 如由于机器故障、人工疏忽等导致记录缺失和输入错误等. 同时, 针对网络拓扑信息未知的前提, 在预处理中需对网络流中出现的 ip 所关联的用户进行识别和定位.

具体步骤如下:

步骤一. 对原始网络数据进行协议解析, 转化成可识别的键值对格式数据.

步骤二. 将网络数据出现的冗余、错误信息, 及属性缺失的数据删除; 删除规则包括:

(1) 网络层报文协议不为 TCP, 作为冗余数据删除;

(2) TCP 报文网络层数据中源、目的 IP 和源、目的端口, 开始时间, 应用层数据中业务类型缺失的, 作为属性缺失数据删除.

步骤三. 将网络数据中多余的属性进行删减. 保留 ID(序号)、STARTTIME(开始时间)、ENDTIME(结束时间)、SRCIP(源 IP)、DSTIP(目的 IP), 实现数据降维, 减少计算复杂度, 提高计算效率, 形成分析数据集.

步骤四. 对网络数据中出现的所有的 IP 地址进行统计, 按照连接数生成 IP 连接分布图, 标记主机用户类型与服务器类型.

步骤五. 在主机用户类型中筛选出连接数很少的主机, 由于连接数未达到一定数量, 无法清晰获得其和其他主机的相似关系, 所以删除此部分的主机, 最后得到主机用户类型的主机集合 U.

2.3 行为序列模式建立

基于数据预处理后获得数据, 基于时间序列, 提取每个用户的行为序列.

序列模式挖掘步骤如下:

步骤一. 根据用户行为序列的定义, 采用字典的方式对用户主机 ip 集合进行编号, 通过遍历主机 ip 集合奖励用户主机 ip 字典.

步骤二. 针对预处理后的分析数据集, 通过每条记录中的 srcip 对数据发送的路径进行序列化处理, 基于用户主机 ip 字典生成每个 ip 用户的访问行为序列;

步骤三. 根据 1.2 节中的最大公共行为子序列计算公式, 得到用户之间的最大公共子序列.

2.4 行为异常分析及预警

由序列模式挖掘得到的结果可以得到用户之间的行为相似度, 并用可视化的方式展现出来, 从而挖掘用户的网络访问行为习惯, 寻找同类的用户之间的共同的访问习惯, 通过比对差异获得异常行为分析结果, 对异常行为进行预警.

行为分析的步骤如下:

步骤一. 取分析数据集前 n 个时间窗数据作为训

练集, 第 n+1 个时间窗作为测试集.

步骤二. 分别求出训练集的所有用户之间的每个时间窗的相似度平均值和方差, 获得每个时间窗的行为相关系数, 从而得到相关系数平均值 RC_{avg} 及相关系数方差 RC_{dx}, 作为检测结果参考. 若测试集的用户之间的相关系数 RC 在 $RC_{avg}\pm RC_{dx}$ 范围内, 则可视为正常用户集合, 反之视为疑似异常用户集合.

步骤三. 对于疑似异常用户集合, 分别比较与其余用户之间的相关系数变化, 若集合中某一用户与多个用户之间都存在相关系数超出参考上下限, 则可判定为异常用户, 加入到异常用户集.

步骤四. 分别将第 1 个时间窗到第 n 个时间窗作为测试集, 其余的作为训练集, 重复步骤 2 和 3. 由此可求出 n+1 个时间窗内每一个时间窗所发生的异常用户和行为.

步骤五. 针对步骤四获得的异常行为, 根据 ip 获取该用户的用户基础属性, 以行为发生的时间作为节点计算用户属性变更系数, 若变更系数为 0 则判定该异常行为产生预警.

3 模拟实验

3.1 实验数据说明

本文采用 ChinaVis 数据集[20]中的流量数据, 采用 python 进行模拟实验. 该数据集包括应用层、网络层和链路层抓取的数据包, 时间跨度为两个月, 共包括约 2000 万条记录. 应用层数据、网络层数据、链路层数据的维度如表 1 所示.

表 1 实验数据格式说明

类型	字段	字段说明
应用层数据	ID	记录序号
	IPSMALLTYPE	IP业务类型, 主要包括66和67, 分别代表了TCP和UDP协议
	FILELEN	网络连接所传输数据的总长度
	FILEAFFIX	网络连接可能传输的文件类型, unk表示未知文件类型
	ISCRACKED	监控系统判断数据包是否有损坏
网络层数据	STARTTIME	开始时间
	SRCIP	源IP
	DSTIP	目的IP
	SRCPORT	源端口
	DSTPORT	目的端口
链路层数据	VPI	虚拟管道标识
	VCI	虚拟管道标识
	ATMAAL1TYPE	虚拟线路中传输数据的属性

3.2 实验结果分析

(1) 分析用户集选取

首先, 在实验数据集中选取时间跨度为一周的数据集合, 基于行为相似度计算选取异常分析的用户集合. 通过识别数据集中的用户形成每个用户的基于时间的行为序列, 序列部分结果如图 4 所示.

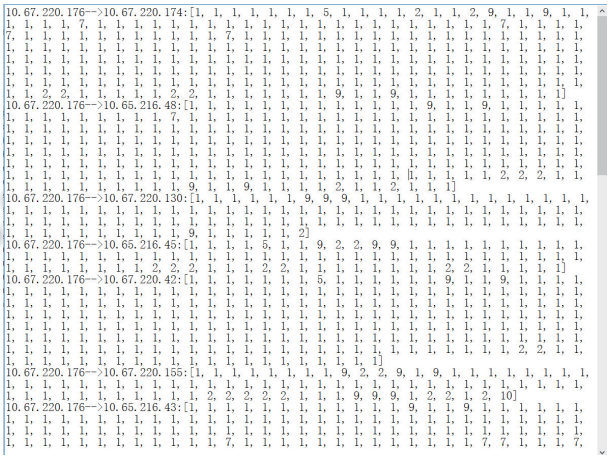


图 4 基于时间的行为序列模式结果

通过用户行为相似度算法生成用户间的行为相似度矩阵. 矩阵中的数值代表两个用户间的行为相似情况, 值越大, 表示行为越相近. 图 5(a) 表示数据集中所有 1000 个用户的 1000×1000 上三角用户行为相似度矩阵, 为验证实验结果, 本文根据数据集中每个 ip 的访问行为发生频度, 并通过对矩阵中相似度值进行统计, 选取其中相似度较高的 6 个用户进行进一步的异常行为分析, 图 5(b) 表示该 6 个用户的 6×6 上三角用户行为相似度矩阵.

由上图所示, 在一周的时间窗内, 该 6 个用户间除了用户 2 和用户 5, 均存在较相似的访问行为.

(2) 行为异常分析

在用户行为序列生成及用户集选取结果基础上, 扩大分析数据的时间跨度, 在实验数据集中选取连续两个月内上述 6 个用户的数据集, 以前 7 周数据作为训练集, 第 8 周作为测试集. 在测试数据集中, 通过系统随机修改一个用户的访问行为序列, 模拟异常行为的发生, 且假设在测试集时间段内分析用户集中用户的基本属性未发生变化, 从而来验证本文的分析方法. 以周为单位计算用户间的相似度, 获得 6 个用户间的两个月内行为相似度曲线分布, 如图 6(a) 所示. 本实验中通过计算两个用户间的行为相关系数平均值和方差判定相似度异常点. 结果如图 6(b) 所示.

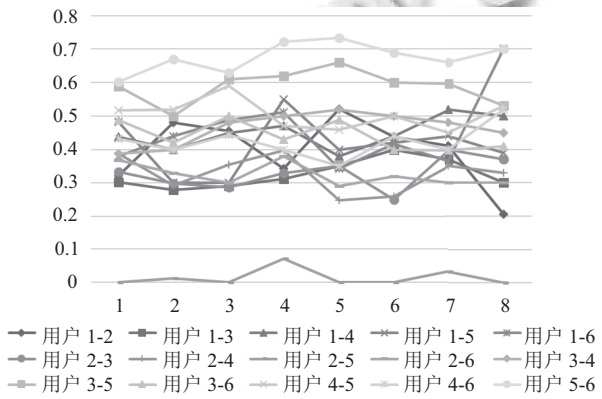
	用户 1	用户 2	用户 3	…	…	用户 998	用户 999	用户 1000
用户 1	1.0000	0.3217	0.3015	…	…	0.4577	0.3729	0.3793
用户 2		1.0000	0.3340	…	…	0.3672	0.4563	0.3674
用户 3			1.0000	…	…	0.3881	0.0000	0.4875
				1.0000	…	…	…	…
					1.0000	…	…	…
用户 998						1.0000	0.5434	0.2342
用户 999							1.0000	0.6025
用户 1000								1.0000

(a) 实验数据集1000个用户行为相似度矩阵

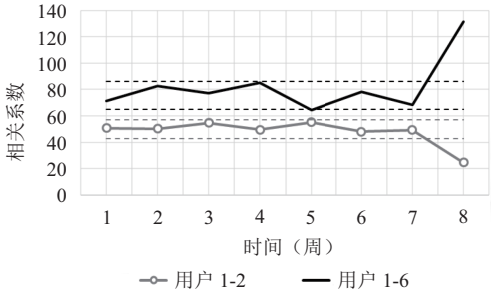
	用户 1	用户 2	用户 3	用户 4	用户 5	用户 6
用户 1	1.0000	0.3217	0.3015	0.4377	0.4829	0.3793
用户 2		1.0000	0.3340	0.3871	0.0000	0.3674
用户 3			1.0000	0.3881	0.5893	0.4875
用户 4				1.0000	0.5177	0.4313
用户 5					1.0000	0.6025
用户 6						1.0000

(b) 实验选取6个用户行为相似度矩阵

图 5 模拟实验用户行为相似度矩阵



(a) 两个月内6个用户的行为相似度分析



(b) 两个月内选取的两组用户行为相关系统比较

图 6 模拟实验用户行为相似度比较

图 6(b) 为分析用户集中用户 1 和用户 2、用户 1 和用户 6 的相关系数分析结果, 根据相关系数计算公式获得用户 1 与用户 2、用户 1 与用户 6 的行为相关系数变化曲线, 并根据训练集数据分别计算其 $RC_{avg}\pm RC_{dx}$. 其中, 浅色线是代表用户 1 与用户 2 的行为相关系数变化, 深色线代表用户 1 与用户 6 的行为相关系数变化, 虚线分别表示两组用户行为相关系数可容忍的偏离上下限. 可见, 浅色线整体比较平缓, 但在第八周突然大幅度下降, 相关系数超出了下限, 表示两个用户的行为差异增大; 同理, 深色线在第八周出现向上突变点, 表示两个用户的行为差异突然减小. 由此推测这 4 个用户中存在异常行为, 由于两组用户中都涉及用户 1, 则可判断该用户行为存在异常. 下一步关联用户的基本属性, 计算属性变更系数. 基于本实验的假设, 用户 1 属性变更系数为 0, 从而判定生成安全预警.

4 结语

就电力企业而言, 内网中同类角色用户的访问网络服务的行为相似度高且变化稳定, 且由于存在较多的协同需求, 在提取单个用户的行为序列后还需要考虑用户间的关联. 本文提出的基于行为相似度的用户行为异常预警方法, 基于行为发生时间建立用户行为序列, 通过用户间行为相似度分析和比对检测异常, 并考虑用户基础属性变更影响, 判定安全预警的生成. 实验数据表明, 在充分训练建立用户行为相似关系后, 能够通过持续监测后续用户行为相似度变化来捕获异常行为. 在后续研究中, 将该方法运用于电力生产环境的网络监测预警系统, 通过实际的网络通信数据集对此模型进行进一步的验证, 并通过逐步增大训练时间窗口, 对方法的误报率等进行进一步的评价. 同时, 在用户行为描述时添加用户访问 URL、访问频度等要素, 使行为特征描述粒度更细, 异常检测结果更准确.

参考文献

1 刘帅. 面向网络数据流的多层面异常行为分析检测技术[硕士学位论文]. 郑州: 解放军信息工程大学, 2015.

2 余勇. 电力系统中的信息安全策略. 信息网络安全, 2003, (9): 31–32.

3 李文武, 游文霞, 王先培. 电力系统信息安全研究综述. 电力系统保护与控制, 2011, 39(10): 140–147. [doi: 10.7667/j.issn.1674-3415.2011.10.026]

4 Thottan M, Ji CY. Anomaly detection in IP networks. IEEE Trans. on Signal Processing, 2003, 51(8): 2191–2204. [doi: 10.1109/TSP.2003.814797]

5 Li L, Lee G. DDoS attack detection and wavelets. Telecommunication Systems, 2005, 28(3-4): 435–451. [doi: 10.1007/s11235-004-5581-0]

6 Kim SS, Reddy ALN, Vannucci M. Detecting traffic anomalies at the source through aggregate analysis of packet header data. Networking 2004. Berlin, Germany. 2003. 1047–1059.

7 梁昇, 肖宗水, 许艳美. 基于统计的网络流量异常检测模型. 计算机工程, 2006, 31(24): 123–125.

8 周颖杰. 基于行为分析的通信网络流量异常检测与关联分析[博士学位论文]. 成都: 电子科技大学, 2013.

9 赵静, 黄厚宽, 田盛丰. 基于隐 Markov 模型的协议异常检测. 计算机研究与发展, 2010, 47(4): 621–627.

10 Das K. Protocol anomaly detection for network-based intrusion detection. GSEC Practical Assignment Version. 2001.

11 秦拯, 李娜, 张大方. Chi-square Distance 在协议异常检测中的应用. 湖南大学学报 (自然科学版), 2005, 32(4): 99–103.

12 连一峰, 戴英侠. 基于模式挖掘的用户行为异常检测. 计算机学报, 2002, 25(3): 325–330.

13 肖喜, 翟起滨, 田新广, 等. 基于 Shell 命令和多阶 Markov 链模型的用户伪装攻击检测. 电子学报, 2011, 39(5): 1199–1204.

14 田新广, 孙春来, 段洣毅, 等. 基于机器学习的用户行为异常检测模型. 计算机工程与应用, 2006, 42(19): 101–103. [doi: 10.3321/j.issn:1002-8331.2006.19.033]

15 陈宁军, 倪桂强, 罗隽, 等. 基于正常行为聚类的卫星通信网异常检测方法. 解放军理工大学学报 (自然科学版), 2008, 9(5): 497–501.

16 郑红艳, 吴照林. 用户行为异常检测模型. 计算机系统应用, 2009, 18(8): 190–192.

17 张炘, 李昆仑. 基于关联规则挖掘的网络行为分析系统设计. 电脑知识与技术, 2011, 7(10): 2333–2334, 2338. [doi: 10.3969/j.issn.1009-3044.2011.10.044]

18 杨铮. 基于流量识别的网络用户行为分析[硕士学位论文]. 重庆: 重庆大学, 2009.

19 Niwattanakul S, Singthongchai J, Naenudorn E, *et al*. Using of jaccard coefficient for keywords similarity. Proc. of the International Multi Conference of Engineers and Computer Scientists. Hong Kong, China. 2013.

20 ChinaVis 2016. http://chinavis.org/2016/challenge.html. [2016].