基于内网用户异常行为安全管理研究

Research on Security Management of Abnormal Behavior of Intranet Users

匡石磊[1]，韦峻峰[2]（1. 中国联通江西分公司，江西 南昌 330096；2. 中国联通河南分公司，河南 郑州 450000）
Kuang Shilei[1], Wei Junfeng[2]（1. China Unicom Jiangxi Branch, Nanchang 330096, China; 2. China Unicom Henan Branch, Zhengzhou 450000, China）

摘　要：
通过提出的SOM-FCM双层聚类模型算法来分析研究企业内部网络用户行为，通过搜集不同模式的日志信息，分析用户的行为特征，将用户的所有行为数据进行聚类，分析聚类后的孤立点和小类来判别其行为是正常还是异常，通过软件工程方式实践该模型算法并在大数据安全平台中使用SOC/4A系统日志数据进行了测试，结果表明该算法可以对内网用户行为数据进行有效地分析与挖掘，能够发现用户违规异常行为及恶意攻击行为。

关键词：
安全管理；日志数据；用户行为
doi：10.12045/j.issn.1007-3043.2019.04.004
中图分类号：TN915.08
文献标识码：A
文章编号：1007-3043（2019）04-0016-05

Abstract：
It analyzes and studies the intranet user behavior through the proposed SOM-FCM double-layer clustering model algorithm. By collecting the log information of different modes, it analyzes the user's behavior characteristics, clusters all the user's behavior data, and analyzes the isolated points and small classes to determine whether their behavior is normal or abnormal. This model algorithm is praticed through the software engineering, and is tested on the big data security platform using SOC/4A system log data. The results show that the algorithm can effectively analyze and mine the behavior data of Intranet users, and users' abnormal behaviors and malicious attacks can be found.

Keywords：
Security management; Log data; User behavior

引用格式：匡石磊，韦峻峰. 基于内网用户异常行为安全管理研究[J]. 邮电设计技术，2019(4)：16-20.

1 概述

随着网络技术的高速发展，数据逐渐升级为企业核心资产，具体表现为数据范围更广、类型更多、规模更大、服务对象更加全面、加工更加深入、管理更为复杂，许多组织（如超市 、银行、电信公司）及一些数据采集系统每日都产生大量的数据，为此各个企业都建立了自己的内部网络。尽管内部网络方便了企业的工作和管理，但内部网络的安全问题频发并且越来越严重，目前有效的安全防护体系已基本形成，但业务逻辑却日趋复杂，安全管理难度也日益增大。企业内部员工很可能在工作时间进行非工作内容的活动或其他异常行为（具体表现为过期账户登录、权限滥用、制度漏洞、异常登录、高频访问、绕行行为），这些都是内部网络需要考虑到的问题。内部网络安全问题的复杂性、不确定性和多样性都会增加解决安全问题的难度。企业内部网络环境中数据威胁通常会造成设备日志粒度粗、日志分散、内容深浅不一，无法对支撑系统进行综合分析，对数据进行关联追踪，原有支撑算法无法满足新业务需要等。因而会产生难以定位实际责任人、内部机密数据泄露、原日志审计难以发现违规、数据分析造成盲区等问题，甚至会影响企业自身的声誉。

为了能够更加有效地检测企业内部网络用户的异常行为，本文对内网络用户的行为进行定义并建立行为模型，明确定义内部用户的行为哪些是异常的，哪些是正常的，再通过软件工程方式验证该模型算法的有效性。本文通过选取企业内SOC系统日志和4A系统日志，分析内部用户的具体行为来检测其是否异常。这对于实际的内网用户的安全检测具有一定的现实意义。

收稿日期：2019-02-22

2 基于内网用户异常行为安全管理设计方案

2.1 系统结构

内网用户行为分析逻辑架构如图1所示。分析层（通过交互查询、策略管理、模型分析、数据合理性检查等分析异常行为、异常登录等问题）将充分利用现有的各类信息日志进行采集、清洗、整合、标签化，构建安全日志的采集层、计算层和存储层，积极实现数据的合理化展示，实现用户的异常行为分析。

2.2 内网用户行为分析流程

内网用户行为分析首先考虑数据来源，实现内网日志数据统一汇聚（见图2），经过数据处理，构建基本的内网行为特征库，根据特征库的相关特征对异常登录行为和异常业务操作行为进行分析，对于异常登录行为，可以通过分析绕过4A系统登录核心业务系统、设备的情况来判断登录的人和登录的IP；而对于异常操作行为，则要通过分析各种异常业务操作行为，挖掘行为人和行为对象来进行判断。在此基础上，分析个人异常行为特征和群体性异常行为特征，挖掘规律性特征。

内网用户行为分析还要考虑数据关联分析，即针对重要业务系统资源，对操作人员和IP进行深层次分析并结合实际业务场景自动发现异常行为，展现审计成果，支撑数据库敏感数据安全管理；用户行为轨迹分析，即实现人员操作行为轨迹分析能力，帮助安全管理员沿着访问行为路径钻取关联信息，挖掘异常点；异常

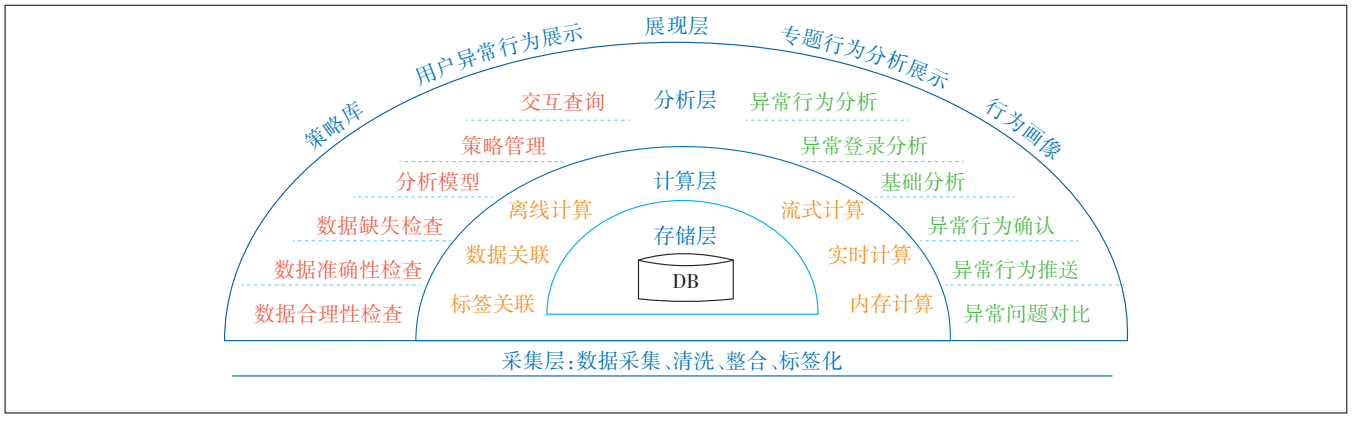

图1　基于内网用户异常行为安全管理技术架构

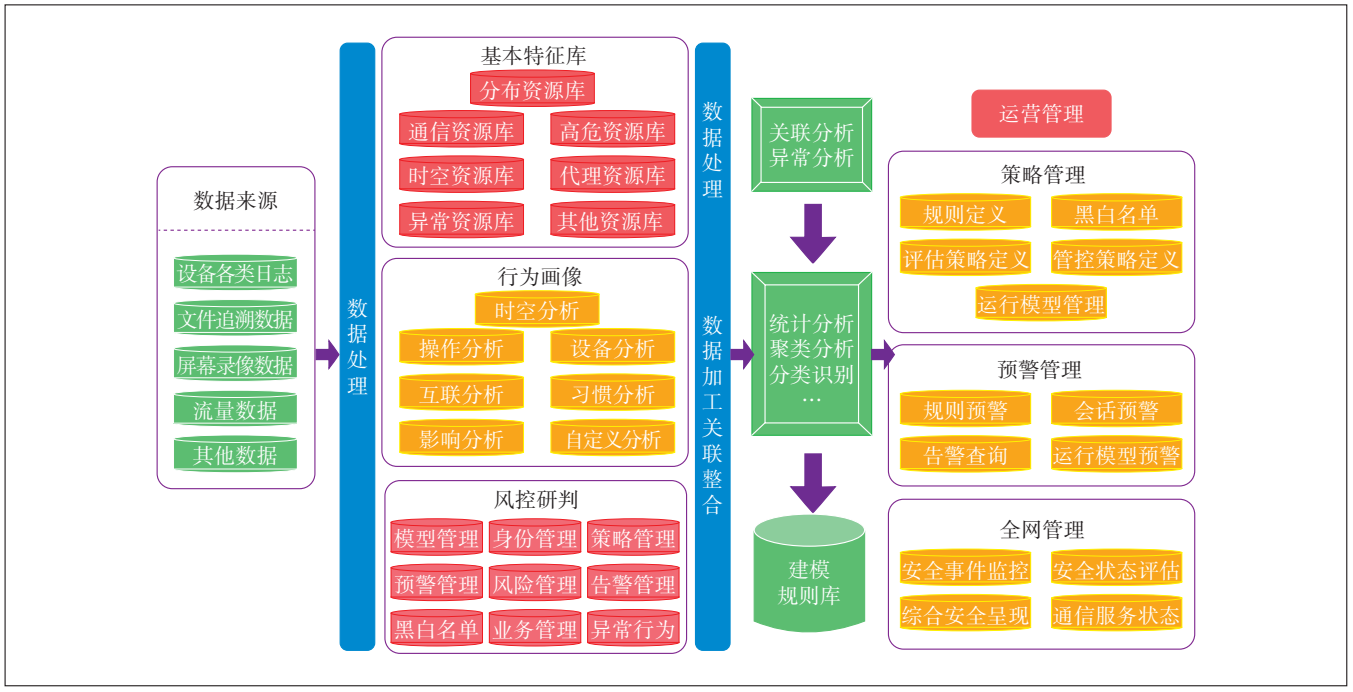

图2　基于内网用户异常行为分析逻辑过程

行为关联审计，即针对高频、高危操作等行为利用基线算法进行关联分析，超过基线的异常访问实时告警，降低人工策略的误差，提高异常行为识别率；疑似攻击行为判断，如账号暴力破解、撞库攻击、非法新账号登录、静默账号突然登录、疑似拖库行为；违规行为判断，如越权操作、同一账号短时间内多个IP登录、闲时段比忙时段操作频次高、敏感信息高频操作。

内网用户行为分析会建立人员画像能力，通过个人特征和群体特征，描绘人员行为的安全风险行为特征，画像能力分为3个规划阶段：个人画像，对个人的操作进行统计，并进行复杂度分析和以往的违规统计；群体画像，统计群体违规行为的违规特征；轨迹特征，从时间维度和空间维度2个方面对违规行为进行相似度分析，挖掘潜在联系。

因此内网用户行为主要工作流程为：

a）汇聚内网日志数据。

b）经过数据预处理构建内网用户行为特征库。

c）通过行为画像、风控研判相关数据加工对数据进一步过滤。

d）利用聚类算法模型并结合大数据技术进行分析。

e）根据聚类检测分析得出异常行为警告及管理分析。

2.3 内网用户行为数据处理

内网用户异常行为检测需要采集内网用户行为信息，根据上述分析过程，内网用户在使用内部网络访问时会产生大量日志（主机、数据库、中间件、网络设备、安全设备等日志），这些日志中包含着大量用户行为记录，把这些日志文件中的数据进行统一的汇聚，如图3所示方式，根据相关标准提取其中的特征，再经过数据预处理，形成内网用户行为特征值来表示用户在这一段时间内的行为。具体采集的日志数据为内网4A/SOC系统中的日志、主机、数据库、中间件和网络设备日志（安全日志已汇聚到了4A/SOC）。

2.4 内网用户行为模型

内网用户行为模型体系的建立主要采用企业内部网络用户常用的SOM-FCM双层聚类模型来分析企业内部网络用户的异常行为。这种分析方法通过搜集不同模式的日志信息，分析用户的行为特征，将用户的所有行为数据进行聚类，分析聚类后的孤立点和小类来判别其行为是正常还是异常。如果出现无法判定的可疑行为，则利用反向选择的法则筛选出正常行为，视其余的行为为异常行为。

SOM-FCM算法，自组织映射学习（SOM—SelfOrganizingMapping）主要通过神经元之间的竞争机制，采用WTA（WinnerTakeAll）获得优胜神经元，建立输入样本与输出之间的映射关系，以达到检测的目的（见图4）。

在竞争阶段，采用最优化匹配原则找到输入向量与权值向量的最佳匹配，如计算欧式距离或取两者内积最大者等，获胜神经元则位于合作神经元的拓扑邻域中心。合作阶段需要确定的是该拓扑邻域的半径，

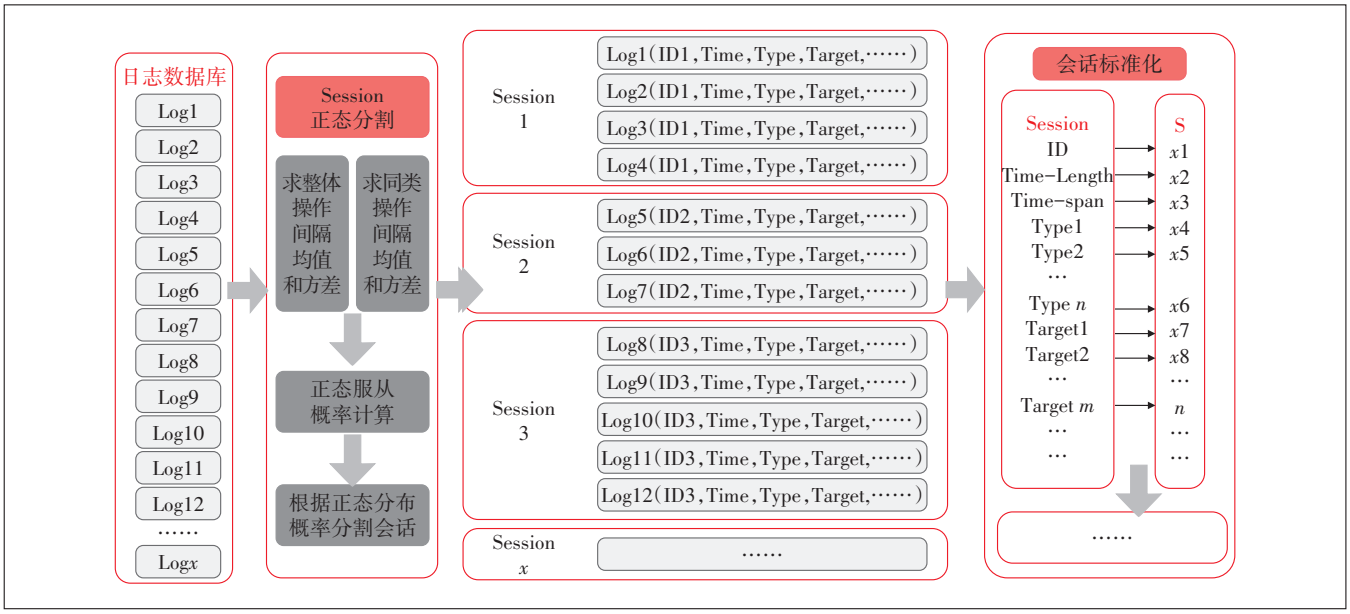

图3 内网用户日志数据处理方式

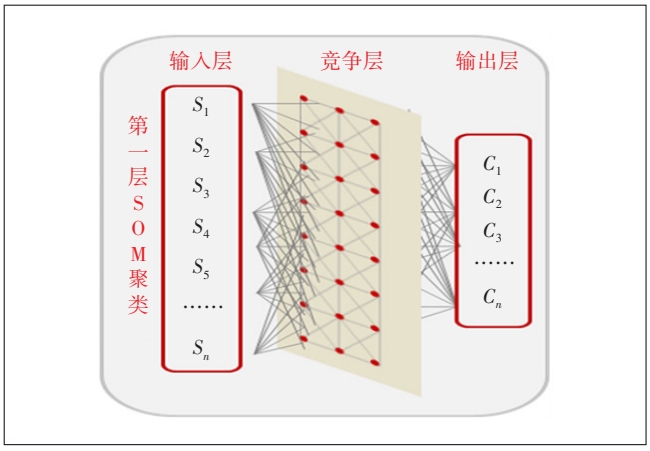

图4　SOM示意图

采用侧向相互抑制的方法能够有效地增强获胜神经元的竞争力。SOM方法适合作为分割集的学习方法，但是，SOM训练不可避免地产生多次迭代，从而增加了训练时间，这限制了其在大规模数据样本尤其是实时性和数据量要求较高的入侵检测系统中的应用。FCM是一种模糊聚类方法。该算法使用隶属度函数计算各个样本点属于某聚类的程度，根据模糊集理论，向量以不同的隶属度归属于不同的聚类集合，按照隶属度的大小来判定样本类别的归属。N个向量被分为C组的聚类结果由下式表示：

$$U=(u_{ij})c\times n \tag{1}$$

式中：

u_{ij}——向量i对分类j的隶属程度

同样计算c个聚类中心，公式如下：

$$c_i=\frac{\sum_{j=1}^{n}u_{ij}^{m}S_j}{\sum_{j=1}^{n}u_{ij}^{m}} \tag{2}$$

将SOM竞争获胜的神经元的权重作为模糊聚类的输入，以实现相似获胜神经元的聚类。FCM通过建立价值函数，并令其最小化实现聚类，建立如下价值函数：

$$(U,C_1\cdots\cdots C_C)=\sum_{i=1}^{N}\sum_{j=1}^{C}u_{ij}^{m}d_{ij}^{2} \tag{3}$$

式中：

m——大于1的实数

u_{ij}——x_i属于j隶属度

d_{ij}——第i个测量到的数据距离类j的聚类中心距离

当价值函数值某个确定的阈值，或它相对上次价值函数值的改变量小于某个阈值，则停止，否则更新隶属度矩阵，重新计算价值函数值。

基于上述双层SOM-FCM算法（见图5）得出数据，再用MaxR-kNN算法会话对比进行分析，其核心思想是如果一个样本在特征空间中的k个最相邻的样本中的大多数属于某一个类别，则该样本也属于这个类别，并具有这个类别上样本的特性，该方法在确定分类决策上只依据最邻近的一个或者几个样本的类别来决定待分样本所属的类别，即kNN方法在类别决策时，只与极少量的相邻样本有关，因此对于类域的交叉或重叠较多的待分样本集来说，kNN方法较其他方法更为适合。在对比分析过程中，首先根据聚类中心和聚类方差确定最大半径，其次由近邻决定新会话分类，最后最大半径决定该会话是否异常，具体如图6所示。

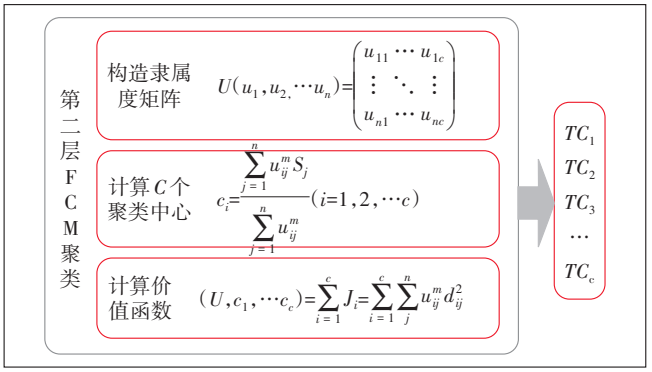

图5　FCM示意图

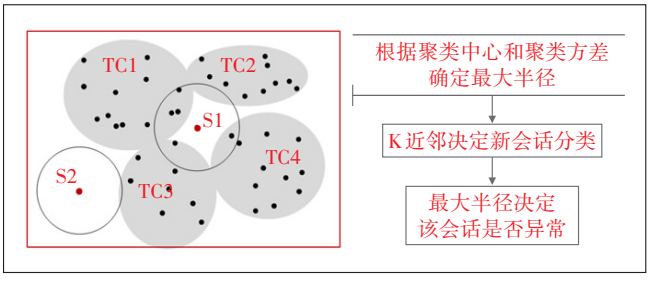

图6　MaxR-KNN示意图

3 内网用户行为实验和结论

3.1 实验环境

实验环境方面：硬件部分采用5台高性能X86服务器，软件部分主机安装redhat7.2操作系统，用户异常行为分析平台基于Hadoop大数据技术架构，融合Spark实时计算、Web2.0可视，使用kafka组件实现主机/数据库日志的海量数据准实时处理，采用数据标准化ETL处理组件对4A系统日志进行处理。本次实验

将传统的分析方法演进为基于数据建模的风险导向分析方法（见图7），利用企业内部4A系统和SOC系统内的登录数据、操作数据等进行标准化分析并通过大数据安全平台系统进行建模挖掘。本次实验重点测试了绕行行为、登录行为、操作行为等不同专题的异常行为分析场景。

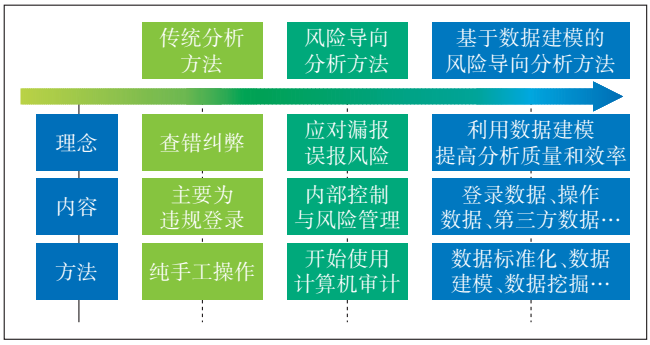

图7　安全数据分析演进

3.2 实验数据和结论

实验的数据集采集自企业内部使用的4A系统日志，该批日志数据中包含了大量内网用户日常工作的各种操作记录，根据实验环境使用的模型定义的正常行为和异常行为，利用模型中的聚类算法对内部网络用户行为进行分类，通过软件工程方式实践后在这些记录数据中发现了存在用户异常行为。

实验环境可以正常呈现内网用户的异常行为，例如异常时间登录、单账号多IP登录、高危操作、敏感数据访问等。

4 结束语

随着信息技术的快速发展，企业内部网络安全管理作为企业无形资产日益受到重视。本文就如何有效地利用企业内部日志数据，提出了建立内网用户行为分析模型，在模型中定义了内网用户的正常行为和异常行为特征，通过双层SOM-FCM聚类算法得出数据，再用MaxR-kNN算法对比进行数据关联得出异常行为检测结果。最后通过软件工程的方式实现方案，通过使用内网SOC/4A系统日志数据进行测试，结果表明能够按角色挖掘对应运维行为，分析出各类运维角色、人员的行为规律、特征属性，分析内网运维场景下特征属性趋势，能够真正对用户异常行为进行有效预警。从而优化授权机制、补齐细粒度管控短板、形成常态化审计机制，做到有效控制运维操作风险以及有效控制业务运行风险。进一步辅助管理层识别、优化运维角色，提供全系统的安全建模依据，最终实现系统的安全稳定运行。

参考文献：

[1] 耿永彬，邢亮，唐国强. 互联网+时代企业网络安全的思考[J]. 通讯世界，2017(1)：58-59.
[2] 周青松. 企业内部网络用户的异常行为分析[D]. 武汉：华中科技大学，2016.
[3] 姜传江，马赟. 企业网络安全结构设计及相关问题解析[J]. 网络安全技术与应用，2016(8)：22-23.
[4] 敖磊，魏煜宸. 企业网络安全架构设计[J]. 网络空间安全，2017，8(z2)：70-72.
[5] 曲佳伟. 企业网络安全防护系统设计与实现[J]. 中小企业管理与科技，2017(30)：164-165.
[6] 冯庆文，李宏宇. 企业网络安全综合防护体系建设[J]. 炼油与化工，2017，28(5)：67-69.
[7] 田广新，孙春来. 基于机器学习的用户行为异常检测模型[J]. 计算机工程与应用，2006，19(2)：432-544.
[8] 施巍巍，邓宏伟. 运营商海量日志分析系统的研究[J]. 江苏通信，2016，32(3)：52-54.
[9] 任凯，邓武，俞琰. 基于大数据技术的网络日志分析系统研究[J]. 现代电子技术，2016，39(2)：39-41.
[10] 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[J]. 解放军理工大学学报：自然科学版，2008(5)：497-501.
[11] 周欢，李广明，张高煜. SOM+K-means2个阶段聚类算法及其应用[J]. 现代电子技术，2010，33(16)：113-116.
[12] 宫改云，高新波，伍忠东. FCM聚类算法中模糊加权指数m的优选方法[J]. 模糊系统与数学，2005，19(1).
[13] 王骁. 基于Hadoop大数据平台资源及用户行为检测技术的研究[D]. 北京：北京交通大学，2015.
[14] 徐时芳，罗晓宾，陈阳华. 基于Spark的分布式大数据分析建模系统的设计与实现[J]. 现代电子技术，2018，41(20)：180-182，186.
[15] 金松昌，方滨兴，杨树强，等. 基于Hadoop的网络安全日志分析系统的设计与实现[C]// 全国计算机安全学术交流会论文集·第2十五卷. 2010.
[16] 应毅，任凯，刘亚军. 基于大数据的网络日志分析技术[J]. 计算机科学，2018，45(S2)：363-365.
[17] 姜传菊. 网络日志分析在网络安全中的作用[J]. 现代图书情报技术，2004(12).

作者简介：
匡石磊，工程师，硕士，主要研究方向为网络安全、云安全等领域；韦峻峰，高级工程师，硕士，主要研究方向为信息安全以及云计算等领域。