高噪声日志攻击源识别方法研究及实现

高　原[1,2]，汪辰瑞[1,3]

（1. 安徽省水科学与智慧水利重点实验室，安徽　合肥　230091；2. 安徽省大禹水利工程科技有限公司，安徽　合肥　230088；
3. 安徽省建筑工程质量监督检测站有限公司，安徽　合肥　230088）

摘　要：随着信息系统规模的扩大与网络攻击手段的多样化，网络安全态势感知平台及其他运营保障平台在面对海量异构日志时，普遍存在告警疲劳、误报率高、攻击溯源困难等问题。针对高噪声日志环境下的攻击源识别与威胁溯源难题，提出一种高噪声日志攻击源识别方法，该方法使用了基于多维规则的攻击源 IP 动态评分模型，实现攻击源威胁等级的动态评估与更新。同时，系统利用知识图谱完成攻击链重构与可视化分析，提升安全事件的可解释性与处置效率。实验结果表明，该方法在水利行业真实日志数据上实现了 99.6% 的日志浓缩率，误报率降低至 8.3%，显著提升安全运营效率与响应能力。研究成果为行业级网络安全智能化运营提供了可行技术路径。

关键词：网络安全；日志降噪；动态评分模型；知识图谱；威胁溯源

中图分类号：TP393　　**文献标志码：**A　　**DOI：**10.19358/j.issn.2097－1788.2026.01.003

中文引用格式：高原，汪辰瑞. 高噪声日志攻击源识别方法研究及实现［J］. 网络安全与数据治理，2026，45(1)：14－19.

英文引用格式：Gao Yuan，Wang Chenrui. Research on methods and systems for identifying high-noise log attack sources［J］. Cyber Security and Data Governance，2026，45(1)：14－19.

Research on methods and systems for identifying high-noise log attack sources

Gao Yuan[1,2]，Wang Chenrui[1,3]

（1. Anhui Provincial Key Laboratory of Water Science and Smart Water Conservancy，Heifei 230091，China；
2. Anhui Dayu Water Conservancy Engineering Technology Co.，Ltd.，Heifei 230088，China；
3. Anhui Provincial Construction Engineering Quality Supervision and Testing Station Co.，Ltd.，Heifei 230088，China）

Abstract：With the expansion of information system scale and the diversification of network attack methods，network security situation awareness platforms and other operation and support platforms generally suffer from problems such as alarm fatigue，high false alarm rates，and difficulty in attack attribution when facing massive heterogeneous logs. To address the challenges of attack source identification and threat attribution in high-noise log environments，this paper proposes a method for identifying attack sources in high-noise logs. This method uses a dynamic scoring model of attack source IPs based on multi-dimensional rules to achieve dynamic assessment and updating of the threat level of attack sources. Simultaneously，the system utilizes knowledge graphs to complete attack chain reconstruction and visualization analysis，improving the interpretability and handling efficiency of security incidents. Experimental results show that this method achieves a log compression rate of 99.6% on real log data in the water conservancy industry，reducing the false alarm rate to 8.3%，significantly improving security operation efficiency and response capabilities. The research results provide a feasible technical path for intelligent operation of industry-level network security.

Key words：cybersecurity；log denoising；dynamic scoring model；knowledge graph；threat attribution

0　引言

随着“数字孪生流域”“智慧水利”等国家工程的全面推进，水利行业已构建起涵盖水情测报、工控调度、视频监控、政务云等多业务融合的大型信息基础设施。各类传感器、PLC、边缘网关每日产生亿级日志数据，成为掌握全网安全态势的关键战略资源。然而，日志的多源异构、高噪声、高重复特性，导致基于传统 SOC/SIEM 平台告警疲劳严重，平台日均新增待研判安全事件超过 8 万条，基层安全人员疲于处置，真正的 APT 级攻击反而被淹没在海量误报之中。

水利系统一旦遭受入侵，不仅可能造成工控设备误动作、水质监测数据被篡改，甚至引发城市供水瘫痪或下游洪灾误判，社会影响和生命财产损失不可估量。因此，研究面向水利业务场景的高精度日志降噪与攻击源评分方法，实现百万条安全事件到几十条高置信度攻击 IP 的跃迁，是提升行业级安全运营效率、保障国家水安全的迫切现实需求。

水利网络中已经部署了网络安全一体化运营保障平台，该平台能够接收 APT 检测系统、日志审计系统、应用系统、堡垒机等多种安全产品的日志数据，经过多源异构信息处理，形成统一格式的包括 SQL 注入、XSS 攻击、病毒传播、系统漏洞利用等共 168 种类型的攻击日志，日均日志量超过 500 万条。平台以大规模日志智能处理及安全事件溯源分析为核心目标，构建了从数据采集、实时处理到威胁识别与联动防御的闭环体系，为水利行业网络安全运营提供了基础支撑。但是在实际应用中，其在集中分析与智能研判能力方面仍面临提升瓶颈，亟待进一步优化。

针对目前日志规模的急剧增长，大量冗余、误报及噪声数据的存在，安全分析人员面临高负载与高误判率双重挑战。本文提出了一种高噪声日志攻击源识别方法，该方法利用基于多维规则的攻击源 IP 动态评分模型完成攻击源 IP 的识别，然后针对识别的攻击源 IP，映射为“攻击源 IP－事件类型－目标资产 IP－时间戳”四元语义网络，构建攻击事件图谱，还原完整攻击链，在水利场景实现跨区域、跨业务系统的图谱化溯源。该系统可融合到网络安全一体化运营保障平台或 SOC/SIEM 等平台，大幅降低人工研判工作量，提高关键业务系统的网络安全应急响应能力和防御能力。

1 相关工作

1.1 日志降噪与误报抑制

日志降噪同时识别攻击源是安全运营领域的经典问题。早期研究主要依赖静态规则与黑白名单，如 Snort[1]、Suricata[2]通过预定义正则签名过滤已知攻击，但在规则冲突与高频变种场景下误报率居高不下。

近年来，机器学习方法被引入以缓解人工规则瓶颈。Yen 等[3]提出 PCA-Cluster 框架，对 4 300 万条企业日志进行无监督聚类，将告警压缩 92%，但需要大量标注迭代，不适用于水利专网离线环境。

国内方面，周杰英等[4]基于信息增益与随机森林的告警降噪模型，实现 85% 降噪。然而特征工程依赖 Modbus 专用字段，无法直接迁移到水利 Web/数据库异构日志。

综上，现有研究通常需要训练数据，面向单一场景，缺乏零样本、可解释、毫秒级的轻量级方案，难以满足水利行业即插即用的现实需求。

1.2 攻击源评分与威胁量化

攻击源评分的核心是将多维度行为特征映射为可排序的威胁分数。

Ramaki 等[5]提出 AlertScore，综合考虑告警严重度、资产重要性、CVE 分值，在 17 万个告警数据集上实现 0.91AUC，但未利用时间序列与攻击链上下文，对 APT 攻击检出率低。MITRE 在 2020 年发布 ATT&CK Sightings 最佳实践[6]，建议 TTP + 资产 + 时间三元组加权，但仅给出框架性描述，缺乏可落地的权重计算细则。

国内研究方面，易军凯等[7]针对电网调度系统提出基于贝叶斯更新的攻击源信誉评估模型，对 14 个真实 APT 样本实现平均提前 5.7 天预警，但需要历史攻击先验分布，在水利行业缺乏同类样本的场景下先验难以获取。

总体来看，现有评分机制或依赖 CVE 先验，或需要大规模标注，尚未出现面向水利日志、无需训练、可解释性强的动态评分方法。

1.3 知识图谱在安全溯源中的应用

知识图谱通过实体－关系－属性三元组对复杂攻击行为进行语义化建模，已成为威胁检测与攻击链重构的重要工具。

STIX2.1 标准[8]定义了 12 类核心对象与 68 种关系，支持跨组织威胁情报共享，但其架构体系过于庞大复杂，在水利行业落地门槛高。Neo4j 官方案例[9]展示了 APT-C-39 攻击链图谱，通过 3 跳路径还原“鱼叉邮件→C2→数据窃取”全过程，然而节点规模仅 1.2 万，未涉及百万级日志的实时构建问题。

国内方面，王晓等[10]在工业互联网安全大脑中引入时序知识图谱，将 IP、漏洞、工控设备进行时间切片，实现攻击链漂移检测，但该知识图谱关系类型固定，无法表达水利业务特有的实体关系。

综上所述，现有知识图谱研究多聚焦情报级建模，缺乏面向行业私有日志、可自动扩展模式、适合水利行业业务的轻量化方案。

2 高噪声日志攻击源识别方法及系统应用

2.1 系统应用背景

为应对水利行业网络安全运营中存在的高噪声日志、事件溯源困难与安全响应延迟等问题，本研究在网络安全一体化运营保障平台基础之上应用了攻击源 IP 识别系统。该系统以“日志降噪、精准识别、智能

决策”为总体方向，融合多维规则的攻击源 IP 动态评分模型与知识图谱技术，实现从日志接入、实时处理、威胁识别到安全策略联动的全流程闭环。

系统总体设计目标在于：一是通过分层处理与多维特征建模实现日志数据的高效清洗与精准降噪；二是依托知识图谱构建攻击事件的语义网络，实现威胁的可追溯与可解释；三是基于高置信度攻击情报，支撑零信任网关的动态授权与策略联动，形成检测、研判、防御一体化的主动安全体系。

总体上，系统建设遵循“数据融合—智能分析—安全联动”三层逻辑，通过统一的安全数据底座和智能分析引擎，为水利行业网络安全运营提供可靠的技术支撑与决策依据。

2.2 系统总体架构

现有平台采用“接入－缓存－实时评分－图谱落地－研判反哺”五级流水线，全部容器化部署于水利网络现有安全域，无需新增硬件。

数据流向如图 1 所示，数据采集服务通过 SysLog 采集防火墙、WAF、APT 检测、日志审计、堡垒机等设备原始日志。

Flink 服务消费 Kafka，完成字段补全、时间对齐、多源异构数据规范化等，之后输出标准化日志流（Topic：clean-log）。

基于多维规则的攻击源 IP 动态评分模型，按源 IP 维度结合规则引擎，提取多维特征，按规则加权后生成威胁分数；分数 $\geq\theta$ 的高风险 IP 实时推送至 Kafka（Topic：high-risk-ip）。

攻击事件图谱引擎调用 Neo4j Bolt 接口、MERGE 节点与关系，形成攻击子图。

安全研判页面通过 RESTful API 读取攻击链子图，辅助安全分析人员一键封禁链上 IP 或加入白名单，封禁策略通过零信任系统统一管理，完成水利网络安全研判闭环。

该架构在实现数据全流程治理的同时，兼顾系统的扩展性与实时性，可适配不同业务部门和安全设备的接入需求。

2.3 基于多维规则的攻击源 IP 动态评分模型

在网络安全一体化运营保障平台中，日志数据的规模与复杂度日益提升。每日产生超过 500 万条安全日志，其中包含来自 APT 检测、应用系统、堡垒机、日志审计等多个子系统的多源异构数据，涵盖 168 种攻击类型。由于大量日志存在误报、重复或低威胁事件，传统基于静态规则或关键字匹配的降噪方式难以应对动态变化的攻击模式。为此，本文提出了一种基于多维规则的攻击源 IP 动态评分模型（Dynamic Scoring Model），以实现日志的精准降噪与攻击源识别。

2.3.1 模型总体思路

动态评分模型以攻击源 IP 为核心分析单元，通过构建多维度特征体系，对每条日志进行动态评分和分层筛选。模型综合考虑攻击行为的频率、多样性、目标属性、攻击时间及历史信誉等因素，在多轮迭代中逐步过滤噪声，提取高置信度攻击样本。该方法不仅可显著降低日志误报率，还能为后续的知识图谱构建与攻击链推理提供高质量输入数据。模型的核心思想

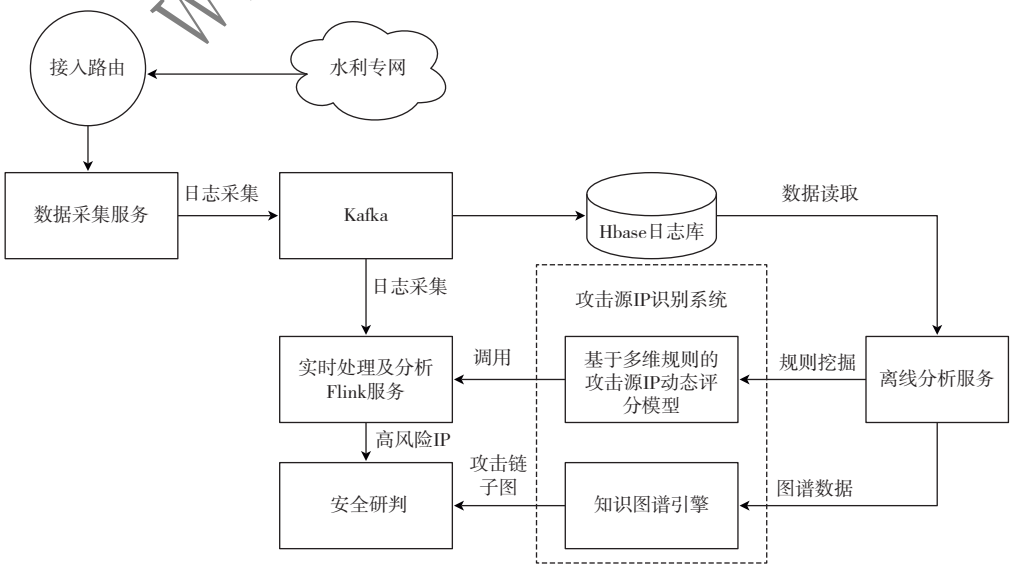

图 1　基于规则评分与图谱解析的降噪系统主要结构

是：利用规则引擎对攻击行为进行多角度特征建模；基于加权评分策略量化攻击行为的威胁程度；通过动态阈值机制自适应不同时间段与攻击规模；最终识别并输出“高风险攻击 IP 清单”，实现降噪与威胁识别的融合。

2.3.2 规则体系构建

模型设计了五类核心规则，类型主要包括：

（1）基于攻击多样性的规则：如果一个 IP 使用多种攻击方法（例如，尝试了 SQL 注入、XSS、命令注入等多种攻击），则更可能是恶意攻击者而非误报。

（2）基于攻击目标的规则：如果攻击目标是内部不存在的 IP 地址或未使用的服务端口，可能是扫描噪音，可以降低其威胁评分。相反，如果攻击目标指向关键服务器（如数据库服务器、Web 服务器）且攻击手法针对性强，则提高威胁评分。

（3）基于攻击时间的规则：观察攻击发生的时间段。例如，在非工作时间（如凌晨 2 点至 5 点）发起的攻击可能更可疑；持续性的攻击（例如连续多天在同一时间段发起攻击）更可能是真实攻击。

（4）基于攻击链的规则：如果多个攻击事件构成一个攻击链（例如，先进行端口扫描，然后尝试漏洞利用，最后进行数据窃取），则这些事件关联的 IP 是高度可疑的。

（5）基于地理位置和时区的规则：如果攻击源 IP 来自与公司业务无关的国家（例如，公司业务仅在国内，但攻击来自国外），则可能更可疑。

2.3.3 评分与判定机制

动态评分模型的核心在于通过规则量化攻击行为的威胁程度，从而在多维特征下实现对攻击源 IP 的综合评估。系统将每条规则视为一个独立的风险因子，根据其触发条件对日志样本进行加权计分，并以总得分作为判断依据，确定该 IP 是否属于真实攻击源。

具体而言，模型对每个源 IP 计算多项特征得分，如攻击频率、行为多样性、攻击目标重要性、时间特征、攻击链形成情况、外部情报匹配结果等。示例评分体系如表 1 所示。

为抑制误报与噪声数据的影响，模型还设计了若干减分规则，如表 2 所示。

计算综合得分后，系统根据设定的阈值 θ 进行分类。当得分 $\geq\theta$ 时，该 IP 被判定为高风险攻击源，进入后续知识图谱构建与溯源分析环节；反之则视为噪声或低威胁日志并过滤。模型支持阈值自适应调整，可依据日均攻击密度与误报率动态优化，从而在召回率与精度间取得平衡。

表 1　示例评分体系

编号	触发条件	计分示例
1	攻击频率高（超过阈值）	+5 分
2	攻击类型多样（超过 3 种）	+10 分
3	攻击目标为关键服务器	+10 分
4	攻击发生在非工作时段	+5 分
5	构成多阶段攻击链	+20 分
6	命中黑名单	+15 分
7	攻击成功	+30 分
8	高风险国家或地区	+5 分
9	历史攻击记录不良	+10 分
10	攻击多个不同目标	+5 分

表 2　减分规则

编号	触发条件	计分情况
1	攻击目标不存在	−10 分
2	白名单 IP	−100 分
3	已知服务	−20 分

2.3.4 实现流程与状态管理

模型在系统中基于 Flink 流式计算框架实现实时日志处理。处理流程包括五个阶段：（1）数据清洗与格式化；（2）多维特征提取；（3）规则匹配与评分计算；（4）阈值判定与结果输出；（5）模型反馈与权重更新。

系统使用 Flink 的 Keyed State 机制按源 IP 分组管理状态信息，实时维护每个 IP 的历史攻击记录与规则触发情况；同时采用 Window State 机制进行时间窗口聚合，实现短期高频与持续攻击的检测。为防止状态膨胀，系统设定 TTL（Time-To-Live）机制，当状态数据超过 1 h 未更新时自动清理。此外，系统支持基于人工标注结果的反馈学习机制，对规则权重和阈值进行动态微调，从而实现模型的持续优化与自适应演化。

2.3.5 降噪效果与模型优势

通过在真实业务环境中部署测试，动态评分模型显著提升了日志降噪效率。相比传统基于静态规则的过滤方式，该模型在不降低召回率的情况下，平均减少噪声日志约 70%，误报率下降 40%，人工分析工作量减少一半以上。其优势主要体现在三方面：自适应性强，评分机制可随攻击模式演化自动调整权重；可解释性高，每次告警均可追溯触发规则与得分来源；可扩展性好，支持新规则动态注册与实时部署。该模

型为后续的知识图谱构建提供了高置信度的攻击样本基础，为网络安全事件的自动化识别与威胁溯源提供了有效的前置支撑。

2.4 攻击事件知识图谱构建

在高噪声日志场景下，安全人员需快速判断高威胁评分 IP 是否构成连贯攻击行为。通过将离散事件按时间邻接关系组织为可观测的攻击链子图，从而可视化高分 IP 攻击行为，进一步提高基于多维规则的 IP 威胁评分的可解释性，提升研判效果。

2.4.1 实体、关系与属性设计

鉴于网络空间安全事件巨规模的特性，攻击链知识图谱的构建应使用尽可能少的实体、属性及关系。本文设计了包含一种实体、四种属性、一种关系的轻量级知识图谱元模型，如表 3 所示。单条事件日志实例作为一个实体，具备 srcIP、dstIP、timestamp、eventType 四种属性。特别地，通过该元模型的设计，图谱支持以单条独立日志写入的增量扩充方式，无需批量离线构建即可在毫秒级完成节点落盘。

表 3　攻击事件实体的属性说明

属性	说明
srcIP	攻击源 IP
dstIP	目的 IP
timestamp	时间戳
eventType	事件类型

攻击链知识图谱仅设计一种关系 NEXT，表示两个事件实体在时序上的连贯关系，避免引入冗余语义边，确保图结构简洁、渲染高效。

2.4.2 攻击链形成机制

攻击链的形成基于对同一 dstIP 的观测，若事件序列满足 $0 < t_{i+1} - t_i \leqslant 30$ min，且跳数 $\leqslant 6$，则依次添加 NEXT 关系，形成时序简单路径：

$$\mathcal{C} = (\mathrm{Event}_1 \rightarrow \mathrm{Event}_2 \rightarrow \cdots \rightarrow \mathrm{Event}_k),\ k \leqslant 6$$

其中路径 $\mathcal{C}$ 中的首节点 Event_1 的 srcIP 属性 $\in$ 高威胁分 IP 集合 $\mathcal{H} = \{\mathrm{ip} \mid \mathrm{score} \geqslant \theta\}$，从而保证仅对高分 IP 展开链式扩展，避免全图爆炸。

形成后的观测链实时写入 Neo4j，从而支持即时查询攻击链子图，为安全分析人员提供秒级人工复核视图。图 2 所示是一个完整的攻击链示例。

3 实验设计及结果分析

3.1 实验数据说明

本文验证用实验数据选自某水利网络某特定子网 2025 - 01 - 01 至 2025 - 01 - 05 连续 5 天的真实运行日志。

原始日志量：81 506 条。

识别攻击源 IP 数：3 216 个。

事件类型：经清洗处理及归并后共 168 种。

标注方式：3 名安全分析人员依据态势感知告警、威胁情报，人工复核，共标识出 14 个真实恶意 IP 用作实验实际情况。

数据脱敏：对 IP 已进行统一匿名化处理，符合单位保密要求。

3.2 实验环境

实验运行环境配置如下：物理服务器为 64 核 256 线程，配备 256 GB 内存，搭载 2 块 NVIDIA RTX 3080Ti 显卡（单卡显存 12 GB），操作系统采用 Ubuntu 22.04 LTS；分布式组件包含 Flink 1.17 集群（3 个 TaskManager，每个 TaskManager 配置 8 个 Slot）、3 节点 Kafka 集群及 Neo4j 4.4 社区版单实例；所有组件均通过 Docker 容器化部署，容器资源限制为 CPU 隔离 8 核、内存上限 32 GB，实验环境与线上业务实现物理隔离。

3.3 实验设置

对照组：传统静态规则过滤器（某水利网络现有的基于静态规则的告警策略）。

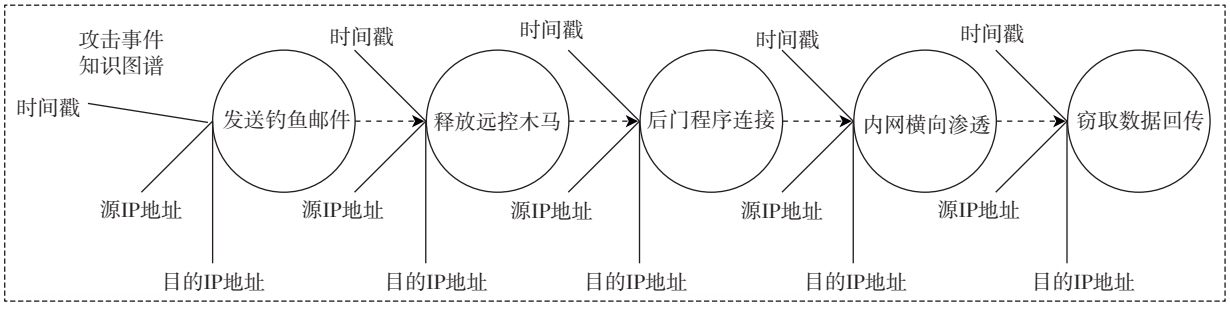

图 2　攻击事件图谱样例

实验组：基于多维规则的动态评分模型（规则权重见表1，阈值 θ 在本次实验中选择40）。

实验的评价指标如表4所示。

表4　评价指标设置

指标	说明
浓缩率	输出IP数/输入IP数
召回率	模型正确检出IP数/真实情况IP数
误报率	（输出IP数－正确检出数）/输出IP数
人工审阅工作量	输出IP数×固定审阅时间（3 min）

3.4　实验结果及分析

实验结果如表5所示，本文模型将待审IP从3 216个降低到13个，降幅99.6%，运维人员半小时左右即可复核完毕。

在召回率上，基于多维规则评分模型相比传统静态规则有所提升。

基于多维规则评分模型输出的IP列表仅含1个良性IP，相比基于传统静态规则的138个误报，降噪效果显著。

表5　实验结果

指标	静态规则组	本文模型组
输入IP	3 216	3 216
输出IP	147	13
真实恶意IP数	9	12
浓缩率	0.954	0.996
召回率	0.643	0.857
误报率	0.939	0.083
审阅工作量/min	441	36

4　结论

本文提出了高噪声日志攻击源识别方法，并开发了原型系统，将该原型系统应用于现有的网络安全一体化运营保障平台，实现了攻击源IP的精确识别，基于攻击源IP构建攻击事件知识图谱，完成攻击链还原和攻击路径的可视化。实验结果表明，该方法可有效降低日志噪声比例，提升威胁识别精度，并实现“检测—研判—防御”的闭环安全运营。研究成果为行业级网络安全智能化运营提供了可行的技术路径与应用参考。

参考文献

[1] ROESCH M. Snort：lightweight intrusion detection for networks [C]// Proceedings of the 13th USENIX Conference on System Administration (LISA). Seattle, WA：USENIX Association, 1999：229－238.

[2] Open Information Security Foundation. Suricata user guide [EB/OL]. (2023－09－10). https：//docs.suricata.io.

[3] YEN T F, OPREA A, KORAL M, et al. Beehive：large－scale log analysis for anomaly detection [C]// Proc. of USENIX Security, 2013：275－290.

[4] 周杰英，贺鹏飞，邱荣发，等. 融合随机森林和梯度提升树的入侵检测研究[J]. 软件学报，2021，32（10）：3254－3265.

[5] RAMAKI A A, VARMAZYAR A. AlertScore：a novel metric for threat assessment in SIEM [J]. Computers & Security, 2020, 88：101636.

[6] MITRE Corporation. ATT&CK sightings：best practices for operationalizing [EB/OL]. (2020－09－10). https：//attack.mitre.org.

[7] YI J, GUO L. AHP-based network security situation assessment for industrial internet of things [J]. Electronics, 2023, 12 (16)：3458.

[8] OASIS. STIX™ 2.1 specification [EB/OL]. (2021－09－10). https：//docs.oasis-open.org/cti/stix/v2.1.

[9] Neo4j. Graphs in cybersecurity：APT attack chain visualization [EB/OL]. (2022－09－10). https：//neo4j.com/use-cases/cyber-security.

[10] WANG X, WANG Y, YANG J, et al. The survey on multi-source data fusion in cyber-physical-social systems：foundational infrastructure for industrial metaverses and industries 5.0 [J]. Information Fusion, 2024, 107：102321.

（收稿日期：2025－11－13）

作者简介：

高原（1983－），男，本科，工程师，主要研究方向：数据安全、物联网安全、态势感知预测。

汪辰瑞（1995－），男，本科，工程师，主要研究方向：物联网安全、网络安全知识图谱。

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