多源安全日志威胁量化分析

冯文英[1], 顾钊铨[1,2], 赵昂霄[1,2], 罗 翠[1], 袁华平[1], 胡 宁[1]

[1]鹏城实验室新型网络研究部 深圳 中国 518055

[2]哈尔滨工业大学(深圳)计算机科学与技术学院 深圳 中国 518055

摘要 入侵检测系统(Intrusion Detection System, IDS)是保障网络安全的重要组成部分, 用于识别和响应恶意活动。IDS 主要依赖预设规则或分类算法来检测异常。基于规则的入侵检测根据预设规则对网络流量进行规则匹配, 但由于其难以自动化适应特定场景, 通常伴随高误报率。基于分类的入侵检测通过机器学习算法对网络流量进行良性或恶意的分类。上述方法难以实现细粒度的威胁程度评估, 因为它们难以处理海量的安全日志并从中挖掘关键信息, 以至于重要的威胁线索和信息维度被忽略。针对这一局限, 本研究创新性地引入回归模型, 提出一种新颖的安全日志威胁量化分析框架 Themis, 以实现对多源安全日志中威胁实体的威胁程度的评估和分析。Themis 首先从多源 Web 安全告警日志中自动抽取出核心威胁实体, 包括安全事件及潜在的恶意 IP 地址。然后设计全面的威胁表示维度, 对抽取出的威胁实体进行多维度特征表示。针对安全日志中普遍存在的标注数据稀少及类别分布不均衡问题, Themis 采用无监督学习技术对威胁样本进行特征增强, 以提升模型的学习效能。最后, 利用增强特征集训练回归评估模型, 进行精细的威胁程度量化和回归分析。通过回归分析, 我们深入探讨并确定了若干对威胁评估具有显著影响的维度; 进一步的消融实验验证了基于特征增强的威胁评估策略的有效性。此外, 系统比较了多种回归算法在威胁评估任务上的性能差异, 同时提供了基于算法效果和复杂度的权衡分析与应用建议。

关键词 入侵检测; 威胁评估; 特征增强; 回归分析

中图法分类号 TP393.1 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2026.03.04

Quantitative Threat Analysis of Multi-source Security Logs

FENG Wenying[1], GU Zhaoquan[1,2], ZHAO Angxiao[1,2], LUO Cui[1], YUAN Huaping[1], HU Ning[1]

[1] Department of New Networks, Pengcheng Laboratory, Shenzhen 518055, China

[2] School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen 518055, China

Abstract Intrusion detection systems (IDS) are critical components of cybersecurity, tasked with identifying and responding to malicious activities. IDS primarily relies on the rules or classification methods to detect anomalies. Rule-based IDSs operate by comparing network traffic against a predefined set of rules to detect anomalies, but they often result in a high false positive rate because they cannot adapt to new scenes. Classification-based IDSs use machine learning algorithms to categorize network traffic as either benign or malicious. These systems often struggle with the granularity required for accurate threat assessment, because the amount of data can overwhelm these systems, leading to important threat indicators being overlooked. To address these limitations, this paper introduces Themis, a novel regression-based framework designed to evaluate and analyze threats present in multi-source security logs. Themis begins by extracting threat entities from web alert logs, which include critical information such as security events and threat IP addresses. These entities are then represented in a multidimensional space, where each dimension corresponds to a specific attribute of the threat entity. To overcome the challenges of data scarcity and class imbalance in security logs, Themis employs unsupervised learning techniques to enhance the features of threat entity samples. The core of Themis is a threat assessment model that leverages these enhanced features to perform threat regression analysis. This model is trained to predict the severity of threats, providing a more precise assessment than traditional intrusion detection methods. To validate the effectiveness of Themis, we conduct detailed regression analysis experiments to explore the dimensions that significantly impact threat severity, as identified through regression analysis. The ablation experiments that demonstrate the benefits of feature-enhanced threat assessment. Furthermore, we compare different regression algorithms used in threat assessment, discussing their respective advantages and disadvantages. Finally, we offer a complexity analysis and practical application recommendations for the various regression algorithms considered.

通讯作者: 顾钊铨, 博士, 教授, Email: guzhaoquan@hit.edu.cn。

本课题得到鹏城实验室重点项目(No. PCL2024A05)、深圳市科技计划(No. KJZD20231023094701003)和国家自然科学基金(No. 62372137)的部分资助。

收稿日期: 2024-07-10; 修改日期: 2024-12-31; 定稿日期: 2026-01-26

Key words intrusion detection; threat assessment; feature enhancement; regression analysis

1 引言

入侵检测是指对系统的运行状态进行监视, 发现各种攻击企图、攻击行为或者攻击结果, 以保证系统资源的机密性、完整性和可用性。随着各类网络攻击不断出现, 针对 Web 层的攻击数量也不断增多。Web 层的安全防护设备包括防火墙、WAF、蜜罐等, 它们不断产生海量的告警数据, 其中既包含真实的安全威胁信息, 也充斥着大量重复、误报与冗余告警。据统计, 在某次体育赛事活动安全保障周期内, 平均每天各类告警设备产生的 Web 告警数之和约 300 万条, 占据空间大小约 9GB。这些告警中仅有极小比例的数据代表了真正的安全威胁。面对如此庞大的数据量, 依赖人工审查不仅耗时费力, 且极易因疏漏而导致关键威胁的忽视。如何从海量告警中识别出重要有价值的告警是入侵检测的关键问题。

实际应用中的入侵检测设备多基于规则, 通过制定规则库对恶意程序或行为进行匹配, 可以有效防御已知攻击, 但对于未知攻击无法识别。在数据量不断增长的情况下, 该类方法存在规则难以制定、规则库难以管理的问题。告警设备根据规则匹配, 仍会产生海量告警, 对于安全人员和安全系统的筛选和研判分析能力仍有很高的要求。基于机器学习分类的入侵检测系统, 在一定程度上提高了检测效率, 但其往往侧重于告警的定性分类而细粒度的威胁评估, 难以有效区分低危与高危告警, 特别是当面临零日攻击或多变攻击策略时, 其准确性和时效性大打折扣。

具体而言, 当前的入侵检测存在如下问题: (1) 缺乏对告警中威胁实体进行威胁量化评估的方法, 因此难以对威胁程度进行直观的研判; (2) 未充分利用海量多源日志间的信息关联进行综合分析, 发现潜在的威胁实体; (3) 对于哪些日志维度在威胁评估中起显著影响作用缺乏深入研究。针对上述问题, 本文提出了一种名为 Themis(Threat assessment of multi-source security logs)的威胁量化评估框架, 用于对 Web 告警中的威胁实体进行威胁程度的评估。首先, 对 Web 应用层的各类日志的类型和格式进行了梳理, 对安全事件类型分布进行初步分析; 然后, 基于日志信息设计威胁评估维度, 对威胁实体样本进行无监督特征增强与表示; 最后, 构建威胁程度回归分析模型, 对威胁实体的威胁程度进行定量评估。本文探究了对于威胁评估影响程度较显著的特征维度, 验证了无监督特征增强对于威胁样本表示的提升作用, 对比了不同的回归算法在威胁评估中的效果差异。

本文的主要贡献为:

(1) 提出了一个覆盖多类型安全日志的威胁量化评估方案。针对 Web 告警日志中的威胁实体(攻击事件及威胁 IP)进行威胁程度评估, 以威胁程度细粒度评估支撑海量告警过滤。

(2) 定义了威胁评估的维度及各维度下的威胁值判定规则。对日志中的安全事件和威胁 IP 基于多源日志中的信息定义评估维度, 并设计各个维度下不同特征取值对应的威胁程度取值, 为威胁量化评估提供了表示框架。

(3) 验证了多源告警日志中对威胁评估具有显著影响的关键特征维度。通过回归分析探究威胁维度对威胁评估的影响表现, 确定在安全运维中应当予以重点关注的重要维度。

(4) 采用了无监督的特征增强方法对威胁实体进行特征表示。对缺少标签数据的威胁样本采用 K-means 聚类确定样本中心, 各样本与聚类中心的距离表示作为增强特征, 提升了对威胁样本的表示和威胁评估效果。

(5) 对比了多种回归算法在威胁评估任务上的性能。实现了基于不同回归算法的威胁评估模型, 分析了不同指标下各算法的效果与复杂度, 为不同场景下的模型选择提供依据。

下文结构安排如下: 第 2 节介绍入侵检测与威胁评估的相关工作; 第 3 节介绍各类型 Web 安全日志的背景知识, 包括日志格式、攻击类型分布等初步分析; 第 4 节介绍威胁评估维度的设计、特征增强方法、威胁评估模型的构建, 及对模型的评价指标; 第 5 节介绍实验, 包括威胁样本的采样与标注规则、威胁评估模型的效果、威胁影响维度探究、特征增强作用验证及不同算法的详细对比; 第 6 节进行本工作的总结与讨论。

2 相关工作

目前进行入侵检测及威胁分析的研究, 主要基于网络流量告警、终端安全日志等威胁数据开展。本节首先介绍基于这两类数据的入侵检测与威胁分析方法, 然后介绍小样本场景下的入侵检测方法。

2.1 基于终端安全日志的入侵检测

终端安全日志包含行为日志、审计日志等多种类型。基于安全日志的入侵检测, 从日志中挖掘潜

在威胁信息, 通过攻击图构建、模式匹配等发现攻击威胁。

审计日志记录了系统、应用程序或用户活动的详细记录, 提供对安全事件可追溯的活动历史。NODOZE[1]使用依赖性图模型和新颖的评分算法对审计日志数据进行实时分析, 高效地重建攻击场景并减少假阳性报警数量。SLEUTH[2]通过将开源情报中的关联与内核审计记录对齐来实时识别和重建攻击链。POIROT[3]基于内核审计记录实时搜索攻击行为并与已知的威胁情报关联, POIROT 采用图形对齐和模式匹配技术, 能快速定位数百万节点规模的日志数据中隐藏的攻击行为, 并有效利用威胁情报中描述的关系来进行威胁狩猎。

行为告警日志记录了各种防护硬件或软件识别的恶意活动产生的信息。AllInfoLog[4]通过四个编码器分别提取语义、参数、时间和其他日志特征的嵌入表示, 然后结合这些嵌入表示来训练基于注意力的双向长短期记忆(Bi-LSTM)模型, 进行异常检测。MMSLAS[5]系统采用分布式架构, 集成业务规则分析、行为分析和基于机器学习的分析方法, 实现对安全日志的深入关联分析和综合处理。MMSLAS 通过高效的日志收集、数据缓冲、流数据处理、多维日志分析、搜索引擎和用户交互等模块, 能够快速定位日志中的异常行为, 并提前检测潜在的恶意请求。ARGUS[6]应用无监督网络表示学习技术来检测大规模日志数据中的攻击。ARGUS 特别针对离散时间图进行优化, 利用历史上下文进行实时检测, 并生成紧凑的攻击场景概览。这些方法通过创建事件链、重建攻击者的行动路径和分析多个事件之间的联系进行安全事件检测与分析。BR-HIDF[7]针对主机入侵检测中的系统调用足迹, 设计了高维空间反特征稀疏的检测框架。

除基于终端日志进行已知攻击的检测外, 许多研究从多类型和多来源的安全数据中全方位挖掘复杂攻击行为, 如多步攻击行为或 APT 攻击[8-11]。ACAM[12]新型网络入侵检测框架基于 MDATA 模型, 对网络安全知识进行高效的表示、管理和利用, 提高了对多步攻击的检测能力。HOLMES[13]基于主机审计数据通过关联可疑信息流来检测 APT 攻击活动。通过关联可能用于实施每个 APT 阶段的攻击技战术, 实时生成高阶图, 以揭示攻击者的步骤。CONAN 系统[14]通过识别 APT 攻击中的三个关键阶段: 部署和执行攻击者代码、收集敏感信息或造成损害、与 C&C 服务器通信或泄露敏感数据, 来集中检测这些阶段并区分恶意行为。

2.2 基于网络流量的入侵检测

基于网络流量的威胁分析通过捕捉网络数据包, 运用相关算法分析数据流, 进行异常检测和威胁探测。

Li 等人提出 HDAN 框架[15], 用于处理多流分类问题, 通过在共同的潜在特征空间中投影来自源域和目标域的数据, 来解决源域和目标域之间的异构性问题。DeepHTTP[16]基于深度学习构建 HTTP 流量检测框架, 不仅能够检测恶意流量, 还能挖掘流量负载中的恶意模式。该框架使用 AT-Bi-LSTM 检测模型, 结合了双向长短期记忆网络和注意力机制, 提高检测的准确性和可解释性。Zang 等人通过语义挖掘对 IP 流量行为进行特征化分析[17], 从端用户的访问关系角度出发, 提取了多个特征来描述流量行为, 并从时间、空间、类别和强度四个维度对行为进行了建模。此外, 针对基于机器学习的网络流量入侵检测规避, Yan 等人研究了在仅有标签的黑盒场景下, 自动规避基于机器学习的网络入侵检测系统的方法[18]。作者提出了一种攻击策略, 不仅自动化所有攻击过程, 而且成功地利用模型提取和转移攻击技术绕过检测。

2.3 小样本无监督场景下的入侵检测

基于终端日志与网络流量的入侵检测均由数据驱动, 需要大规模采集威胁数据, 依托庞大的计算资源训练。而实际应用中, 物联网节点的资源受限, 导致模型训练只能使用小规模数据, 从而降低了检测率。

Diallo 等人提出了一种新颖的基于自适应聚类的入侵检测系统 ACID[19], 用于在网络边缘进行恶意流量分类。ACID 利用轻量级神经模型学习低维嵌入, 以解决传统基于规则的网络入侵检测系统在应对新型和复杂攻击时的不足。Wei 等人提出了一种基于特征增强的恶意流量检测模型 FE-MTDM[20], 用于检测小规模不平衡数据集中的网络威胁。该模型通过聚类算法增强原始流量特征, 并结合浅层神经网络和随机森林算法构建了一个双分类模型。TMG-IDS 使用基于生成对抗网络的数据增强进行网络入侵检测[21]。

综合以上分析, 目前的入侵检测与威胁分析多为基于有监督的机器学习分类方法, 尽管在已知攻击数据集上达到较高的检测准确率, 但难以实现对威胁实体的细粒度威胁程度评估。

3 基础背景

安全日志主要包括两类: 主动威胁探测与安全设备告警。本研究中, 选择蜜罐日志代表主动威胁探测数据, 选择 WAF 日志代表传统安全设备告警数据,

进行 Web 应用层的威胁量化分析。本部分首先介绍蜜罐和 WAF 的基本概念、然后介绍蜜罐日志和 WAF 日志的格式, 最后对其重要字段的取值分布进行了初步统计分析。为便于后续描述, 在本文中对如下概念进行定义。

定义 1. 威胁实体: 安全日志中出现的可对被保护资产产生危害的主体。在本研究中, 具体指安全事件和威胁 IP。

定义 2. 安全事件: 安全日志中的一条记录代表一个威胁事件, 包括踩蜜事件和 WAF 告警事件[1]。在本研究中, 踩蜜事件指蜜罐日志中的一条记录, WAF 告警事件指 WAF 日志中的一条记录。

定义 3. 威胁 IP: 威胁发起的物理主体, 如无进一步分析, 一般指告警记录中的源 IP。此外, 还包括 HTTP 数据包 payload 中可能包含的 IP。

3.1 日志类型介绍

本部分介绍蜜罐和 WAF 的功能及其日志包含的信息。

3.1.1 蜜罐日志

蜜罐是一种存在价值即为被探测、被攻击和被攻陷的安全资源, 主要目的是欺骗防御[22]。蜜罐没有业务上的实际用途。其所有流入或流出蜜罐的流量都预示着扫描、攻击及攻陷可能性, 用以监视、检测和分析攻击。

根据蜜罐的部署目标不同, 蜜罐分类为产品型和研究型[23]。产品型侧重于分流真实的网络攻击流量, 吸引攻击者把注意力和目标从真实系统转移到蜜罐, 常见于非安全业务为主的公司和网络。研究型蜜罐主要是由安全公司和安全研究人员部署, 主要目的是收集攻击流量, 研究攻击行为、了解攻击意图和提取攻击主体特征等, 基于蜜罐获得第一手、最真实的网络攻击数据并应用于安全防御产品的改进升级和测试。

蜜罐日志在蜜罐系统中记录关于网络活动、攻击行为、漏洞利用尝试以及其他安全事件。蜜罐日志记录的信息通常包括: (1) 攻击信息。攻击者对蜜罐系统的攻击尝试的详细信息, 包括攻击类型、攻击载荷和攻击者的 IP 地址等。(2) 失陷信息。当攻击者成功入侵蜜罐系统时, 失陷日志会记录攻击者在系统内的活动, 包括所访问的文件和命令执行等。(3) 样本信息。样本日志包含了攻击者上传到蜜罐系统的恶意文件样本。

3.1.2 WAF 日志

WAF(Web Application Firewall)即 Web 应用防火墙。WAF 部署在 Web 应用程序前, 在用户请求到达 Web 服务器前对用户请求进行扫描和过滤, 分析并校验每个用户请求的网络包, 确保每个用户请求有效且安全, 对无效或有攻击行为的请求进行记录或隔离。WAF 部署在网络边界, 可以保护 Web 应用程序免受恶意入侵和攻击, 如 SQL 注入、跨站脚本(XSS)攻击、DDoS 攻击等。

WAF 日志是指由 WAF 生成的记录文件, 用于追踪和记录所有经过 WAF 处理的网络请求及其相关安全事件。WAF 日志通常包括请求时间、客户端 IP、请求方法和 URL、请求头和响应头、规则匹配情况等信息。

3.2 日志格式介绍

本文附录中的表 1 和表 2 展示了典型的蜜罐日志和 WAF 日志中包含的字段及其介绍。蜜罐日志字段主要有: 踩蜜事件标识“_id”、事件发生时间“time”、源 IP“src”与源端口“sport”、目的 IP“dst”与目的端口“dport”、访问协议“protocol”、“action”、踩蜜行为类型与描述“attack”、“desc”、“behavior”、踩蜜行为的风险等级“threat_level”与“risk”等。

WAF 日志字段主要有: 告警标识“_id”、用户设备信息“uaAgent”、访问域名“domain”、目标地址“url”、告警发生时间“time”与接收时间“received_time”、源 IP“src”与源端口“sport”、目的 IP“dst”与目的端口“dport”、告警行为描述“desc”、客户端 IP 的地理位置“location”与经纬度“longitude”、“latitude”、运营商信息“operators”等。每一条蜜罐或 WAF 日志记录代表客户端 IP 对蜜罐或 WAF 的一次访问事件。不同厂商的蜜罐日志格式有所不同, 但基本功能和字段格式大致相同, 本文以本研究支撑项目中的某种典型 json 格式的蜜罐和 WAF 日志进行研究。

3.3 日志初步解析

为了解实际发生的攻击种类与分布情况, 我们对某安全周期内的蜜罐和 WAF 记录到的安全日志进行了初步分析。在该安全周期内, 蜜罐产生约 200 万条数据, 平均每天产生踩蜜记录 40 万条, 平均每秒产生踩蜜记录 5 条; WAF 产生 1300 万条数据, 平均每天产生告警记录约 260 万条, 平均每秒产生约 30 个告警。同样时间内, WAF 产生的告警约为蜜罐记录

① 本研究中的安全事件包括踩蜜行为和 WAF 告警, 踩蜜行为本身不属于攻击事件, 但提示了存在攻击风险。WAF 告警代表了存在攻击, 但有些是没有实时性威胁的扫描、误报或者重复告警。

数量的 6 倍。

图 1 和图 2 展示了该安全周期内蜜罐日志和 WAF 日志中的攻击类型统计。在蜜罐日志中, 多数踩蜜行为普通访问行为(蓝色), 源 IP 仅对蜜罐建立访问连接。仅有约 1/6 的踩蜜记录在踩蜜之后还进行了其他的行为(非蓝色的五项), 如登录尝试、爬虫、触发系统绊线等。在 WAF 日志中, “CC 策略防护”类告警最多, 约占总告警数的一半; 其次为“信息防泄露”, 约占 1/4; 其他告警的攻击类型包括: 尝试 SYN 握手、HTTP GET 请求等。从上述初步分析看出, 各攻击类型的数量分布并不均匀, 并与设备规则定义强相关, 应对某些特定高危踩蜜或告警类型如“触发系统绊线”“注入攻击”等进行重点关注和分析。

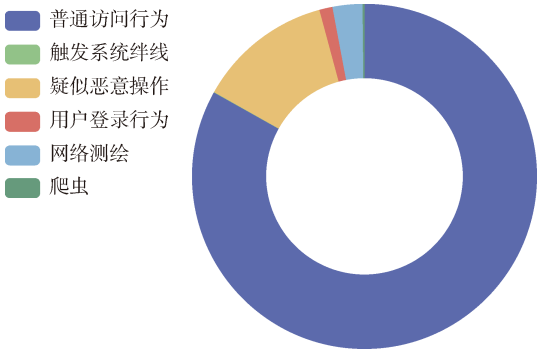


图 1 踩蜜事件类型统计

Fugure 1 Statistics of honey stepping event types

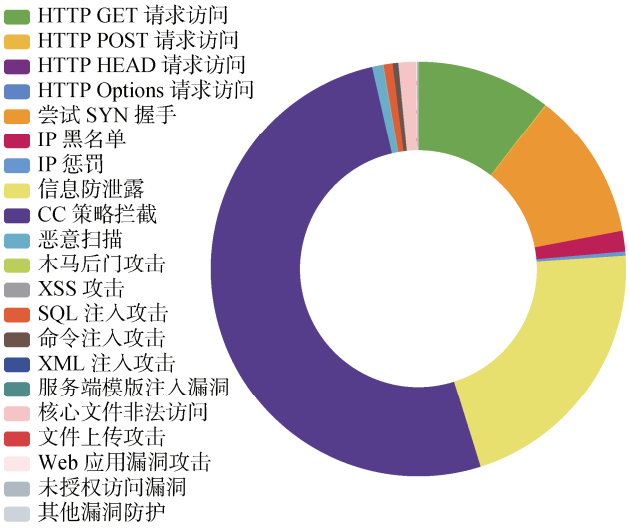


图 2 WAF 告警事件类型统计

Figure 2 Statistics of WAF alarm event types

4 威胁量化评估

本研究基于多源威胁日志提出如图 3 所示的威胁评估框架 Themis, 针对威胁事件和威胁 IP 设计了不同的评估维度, 运用不同回归算法进行回归分析[24-25]。威胁评估流程自底向上。Themis 接入的日志类型包括: 蜜罐日志、WAF 日志、其他日志。数据接入后, 经过如下的模块完成威胁评估:

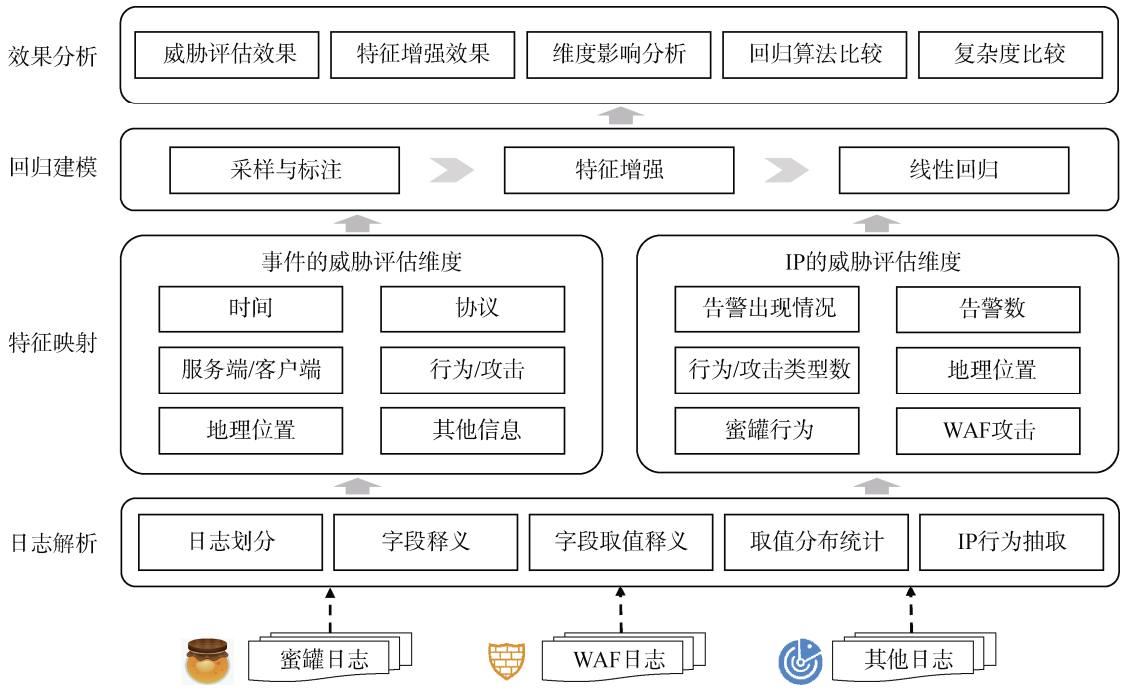


图 3 Themis: 多源日志威胁量化评估框架(橙色标注模块依赖一次性人工工作, 模型迁移和精度提升则需要多次人工参与迭代提升)

Figure 3 Themis: Quantitative threat assessment framework for multi-source security logs

(1) **日志解析:** 对接入的日志进行划分、字段释义、字段取值统计及 IP 行为抽取。

(2) **特征映射:** 对威胁实体基于设计的维度进行特征表示。对安全事件的评估维度包括时间、协议信息、客户端/服务端信息、行为信息、地理位置、其他信息。对 IP 的评估维度包括: 在不同告警设备中的出现情况、告警数量、行为或攻击类型数量、地理位置、踩蜜行为类型、WAF 告警行为类型。将各类日志中的每条记录根据各字段的取值映射为不同维度的特征, 作为一个样本表示。

(3) **回归建模:** 对日志进行抽样与标注、特征映射和无监督条件下的特征增强之后, 用回归算法进行建模。

(4) **威胁评估与效果分析:** 在实际部署中即可用构建好的回归模型进行威胁程度的计算与评估。本研究中在模型构建完成后, 详细分析了回归威胁建模的效果、各维度对威胁评估结果的影响等。

4.1 威胁评估维度设计

本研究基于各类型日志中包含的字段信息, 挑选出与威胁评估相关的信息字段, 据此设计评估维度。表 1~表 3 介绍对安全事件和威胁 IP 评估的各维度、对应采用的日志字段、字段取值及威胁设定分数。本研究中, 对威胁程度的评分范围为从 1~10, 分数越高代表威胁程度越高。

4.1.1 踩蜜事件威胁评估维度

表 1 列出了对踩蜜事件进行评估的维度(“维度”列), 每个维度对应的日志字段(“字段”列), 每个维度/字段下的取值(“取值”列)及我们为其设定的威胁得分(“威胁分值”列)。对各维度、字段、取值的威胁分值设定规则如下。

- **时间信息:** 若安全事件发生在被保护周期内的敏感时间段(如赛事活动举办期间), 则视为具有更高的威胁值, 设为 10; 若发生在其他时间, 则认为威胁值较低, 设为 5。
- **服务端信息:** 包括被访问的域名和端口。若被访问的域名为被保护范围的关键资产, 则设定较高的威胁值 10, 否则设定较低的威胁值 5。若被访问的端口为 80 端口(http 协议默认端口), 威胁值设为 6; 若端口号为其他, 代表可能存在异常访问行为, 威胁分值设为 10。
- **协议信息:** 蜜罐可以记录到踩蜜 IP 发起连接的协议。若访问行为处于建立 TCP 连接阶段, 威胁值设为 6; 已建立 HTTP 连接后, HTTP GET 和 HTTP HEAD 连接的威胁值设为 6, HTTP POST 用于向指定资源提交数据, 可能会导致新的资源创建或已有资源的修改, 威胁值设为 8, HTTP OPTIONS 用于获取目标资源所支持的 HTTP 请求方法, 客户端可能利用该方法来测试服务器的性能, 威胁值设为 8。
- **行为信息:** 踩蜜行为共包括 6 种类型。根据不同行为设定不同的威胁分数: “普通访问行为”仅对蜜罐进行了基本访问, 威胁值设为 5; “触发系统绊线”代表潜在攻击者在蜜罐系统内部进行了一系列活动, 威胁程度较高, 威胁值设为 10; “网络测绘”威胁程度较低, 威胁值设为 6; “用户登录行为”也代表了攻击者可能存在攻击意图, 威胁值设为 8; “疑似恶意操作”为高危操作, 威胁值设为 10; 爬虫行为威胁程度设为 8。
- **风险等级:** 蜜罐日志一般给出该条记录的粗粒度风险级别, 分为低中高三档, 分别赋予 3、6、9 的威胁分值。

表 1 踩蜜事件威胁评估维度

Table 1 Dimensions of threat assessment for honey stepping events

维度	字段	取值	威胁分值
时间信息	time	敏感时间段	10
		其他时间	5
服务端信息	dst	业务资产相关域名	10
		其他	5
	dport	80 端口	6
		其他端口	10
协议信息	protocol/action/attack	TCP	6
		HTTP_ GET	6
		HTTP_HEAD	6
		HTTP_POST	8
		HTTP_ OPTIONS	8
行为信息	desc/behavior/honeyAction	普通访问行为	5
		触发系统绊线	10
		网络测绘	6
		用户登录行为	8
		疑似恶意操作	10
		爬虫	8
风险等级	threat_level/risk	低	3
		中	6
		高	9

4.1.2 WAF 告警事件威胁评估维度

表 2 列出了对 WAF 告警事件进行评估的维度(“维度”列), 每个维度对应的日志字段(“字段”列), 每个维度/字段下的取值(“取值”列)及我们为其设定的威胁得分(“威胁分值”列)。对各维度、字段、取值的威胁分值设定规则如下:

表 2　WAF 告警事件威胁评估维度

Table 2　Dimensions of threat assessment for WAF alarm events

维度	字段	取值	威胁分值
时间信息	time received_time	敏感时间段	10
		其他时间	5
		延迟小于 1500ms	4
		延迟大于等于 1500ms	8
客户端信息	uaAgent	包含 OPPO、SAMSUNG、Redmi、HarmonyOS、HUAWEI 等常见手机型号	4
		其他	8
服务端信息	domain/dst url dport	dst 为域名	10
		dst 为其他	5
		含有字符串“passwd”	10
		其他	6
		80 端口	6
		非 80 端口	10
攻击描述信息	desc	HTTP GET 请求访问	
		HTTP POST 请求访问	2
		HTTP HEAD 请求访问	2
		HTTP OPTIONS 请求访问	2
		尝试 SYN 握手	2
		IP 黑名单	2
		IP 惩罚	6
		信息防泄露	6
		CC 策略拦截	6
		恶意扫描	6
		木马后门攻击	8
		XSS 攻击	10
		SQL 注入攻击	10
		命令注入攻击	10
		XML 注入攻击	10
		服务端模板注入漏洞	10
		核心文件非法访问	10
		文件上传攻击	10
		Web 应用漏洞攻击	8
		未授权访问漏洞	8
		其他漏洞防护	8
位置信息	location	srcIP 地址为业务区域	5
		srcIP 地址为其他地区	10
运营商信息	operators	国内三大运营商	6
		其他	8

- **时间信息：**时间信息包含告警发生时间(time)和接收时间(received_time)两个字段的特征。告警发生时间威胁值的设定规则同蜜罐的“time”字段一样。接收时间代表 WAF 设备接收到告警的时间，若其与告警发生时刻的延迟不超过 1500 ms，则威胁值较低，设为 4，否则设定较高的威胁值 8。
- **客户端信息：**客户端信息为 User Agent(用户代理)，是指浏览器或其他客户端发送给服务器的包含有关客户端软件、操作系统、浏览器等信息的 HTTP header 字段。若客户端包含常用手机型号，有可能为业务人员或运维人员的访问导致的误报，设定较低的威胁值 4。若为其他，设定较高的威胁值 8。
- **服务端信息：**除包含目标域名和端口，还包含访问资源的 url。目标域名和端口的威胁值设定同蜜罐日志。对于 url，若包含敏感字段，如“passwd”，则设定最高的威胁值 10，否则设定 6。
- **攻击描述信息：**攻击描述信息代表了 WAF 设备判定的告警所属类型。本研究将 WAF 中出现的攻击分为四类，详细介绍如下，对各攻击类型的威胁分值设定见表 2。

① **协议行为：**HTTP GET 请求访问：客户端向服务器请求信息；HTTP POST 请求访问：向服务器发送数据；HTTP HEAD 请求访问：向服务器发送数据；HTTP OPTIONS 请求访问：获取特定 URL 支持的所有 HTTP 方法；尝试 SYN 握手：建立 TCP 连接。

② **安全策略下的违规行为：**IP 黑名单：阻止特定 IP 地址或 IP 地址范围的网络访问；IP 惩罚：临时或永久性限制特定 IP 地址访问服务器的措施；信息防泄露：防止组织内部敏感信息未经授权就被泄露出去的安全策略；CC 策略拦截：CC 策略拦截是针对 CC 攻击(Challenge Collapsar，DDoS 攻击的一种)的防御手段，通过监测并限制异常流量、识别和阻断恶意请求，避免服务器因处理过多无效请求而无法正常服务。

③ **特定类型攻击行为：**恶意扫描：使用自动化工具对目标网络或主机进行有目的性的探测和分析；木马后门攻击：通过在目标系统中植入木马程序，创建隐蔽通道，以便随时无需授权即可远程访问和控制该系统；XSS 攻击：跨站脚本攻击；SQL 注入攻击：插入恶意 SQL 语句片段；命令注入攻击：恶意构造输入；XML 注入攻击：构造恶意 XML 数据，利用程序解析 XML 数据时存在的漏洞，注入和执行恶意 XML 命令或代码；服务端模板注入漏洞：针对动态网页生成中的模板引擎，攻击者通过提交恶意数据，使得模板引擎在渲染页面时执行了非预期的服务器端代码；核心文件非法访问：未经授权访问或操作系统的内核文件、配置文件或关键系统文件；文件上传攻击：上传恶意文件。

④ **特定漏洞下的攻击行为：**Web 应用漏洞攻击：借助 Web 应用程序中存在的各种漏洞发起攻击；未授权访问漏洞：无需任何身份验证或凭据便能访问或操作受保护资源的漏洞；其他漏洞防护：一系列

针对各类未知或已知漏洞的预防措施。

- **其他信息:** 包括位置信息和运营商信息。若告警的 srcIP 地理位置为业务区域, 则可能为误报, 设定较低威胁值; 其他地理位置的 IP 告警则设置较高威胁值。运营商的威胁值设定同理, 若为本地常见运营商或业务系统合作运营商, 威胁值设定较低, 其他则设定较高的威胁值。

4.1.3 IP 威胁评估维度

表 3 列出了对威胁 IP 进行评估的维度(“维度”列), 每个维度对应的取值(“取值”列)及我们为其设定的威胁得分(“威胁分值”列)。对各维度、字段、取值的威胁分值设定规则如下:

表 3 威胁 IP 评估维度

Table 3 Assessment Dimensions of Threat IP

维度	取值	威胁分值
在各设备中出现情况	仅出现在蜜罐	5
	仅出现在 WAF	5
	同时出现在蜜罐和 WAF	10
行为告警数(包括蜜罐和 WAF)	< 5	2
	5~25	4
	25~50	6
	50~100	8
	100 以上	10
攻击种类数量	仅发起一种行为	5
	发起了 2 种行为	8
	发起了 3 种及以上	10
蜜罐行为类型*	蜜罐行为 behavior 评分	见表 1
WAF 攻击类型*	WAF desc 评分	见表 2
地理位置	srcIP 地址为业务区域	5
	srcIP 地址为其他地区	10

(*注: 蜜罐行为信息只有出现在蜜罐中的IP才有, WAF 攻击类型只有出现在 WAF 告警日志中的 IP 才有。若同一 IP 被记录到多条踩蜜行为信息或多条 WAF 告警信息, 则取其威胁分值最高的行为记录。)

- **在各类型日志中的出现情况:** 若威胁 IP 仅在蜜罐日志中或 WAF 日志中出现, 威胁程度较低设为 5。若同时出现在所有的日志类型中, 威胁值设置为较高的 10 。
- **行为告警数:** 威胁 IP 在各个日志中出现的总告警次数, 在本研究中为踩蜜次数与 WAF 告警数之和。根据告警次数的多少, 设定不同的威胁分值。
- **行为或攻击种类数:** 威胁 IP 的行为类型数。若威胁 IP 仅进行了一种行为, 则威胁值较低, 行为种类数越多, 威胁值越高。
- **蜜罐行为信息:** 根据威胁 IP 踩蜜行为类型设定威胁程度, 威胁分值评定规则见表 1。
- **WAF 攻击信息:** 根据威胁 IP WAF 告警类型设定威胁程度, 威胁分值评定规则见表 2。
- **地理位置信息:** 根据威胁 IP 所处的地理位置设定威胁值, 威胁分值评定规则见表 2。

4.2 威胁实体表示与特征增强

依据上述设计的维度, 可以对安全事件和威胁 IP 进行特征表示。本部分介绍样本的特征表示与增强, 对威胁评估问题进行定义, 然后介绍对威胁实体的建模原理。

对安全日志中威胁实体的威胁程度进行人工标注需要花费较大的人力, 因此威胁评估面临样本标签稀缺问题。此外, 各威胁维度下的威胁值分布存在不均衡情况。针对小样本不均衡场景下的威胁评估, 有研究采用基于聚类中心的无监督方法, 用于丰富样本特征, 增强模型的表示能力, 在小样本不均衡的流量样本分类任务中相比于不加入特征增强的分类准确率有所提升[19]。

本文采用基于聚类中心的特征增强方法, 增强无标签数据中样本的丰富度。如图 4 所示, 灰色圆点代表样本空间中的威胁实体, 红色圆圈代表所有实体的聚类中心。计算每个样本到聚类中心的距离, 将该距离作为增强特征, 加入威胁实体的原始维度下的特征表示中。整合后的样本特征表示作为该威胁实体的样本表示, 输入威胁评估模型中。

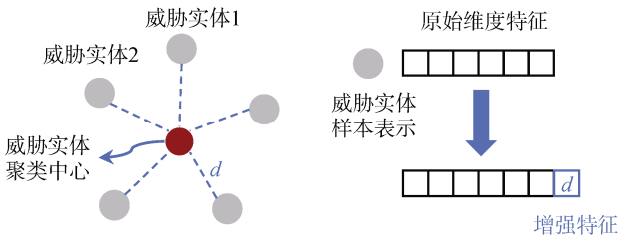


图 4 特征增强示意图

Figure 4 Schematic diagram of feature enhancement

若某条安全事件的蜜罐日志的文本表示如图 5 所示。根据 4.1.1 节介绍的各维度评分规则, 得到其各维度的原始特征表示为(10, 10, 6, 6, 5, 3)。经过聚类, 计算出该样本距离聚类中心的距离为 5.16, 则加入特征增强后的样本表示为(10, 10, 6, 6, 5, 3, 5.16)。对 WAF 告警事件的表示同理。若某威胁 IP 各维度特征为: 同时出现在蜜罐和 WAF 日志中, 行为告警数共计 23, 踩蜜和攻击行为种类数共计 3 种, 蜜罐行为为“疑似恶意操作”, WAF 攻击为“CC 策略拦截”, 地理位置为非业务区域。该样本距离所有威胁 IP 的聚类中心的距离为 3.28。依照 4.1.3 节的定义, 得到其特征表示为(10, 4, 10, 10, 6, 10, 3.28)。

4.3 威胁评估回归模型

本研究中, 模型的输入样本(踩蜜事件、告警事

件或威胁 IP)表示为 X, 事件或 IP 的最终威胁评分作为标签 y。对其进行威胁量化评估的问题定义如下:

```
{
    "time": 1697626392816(2023-10-18 18:53:12),
    "dst": "achi****.****fair.org.cn",
    "protocol": "TCP",
    "dport": "80",
    "attack": "尝试 SYN 握手",
    "behavior": "普通访问行为",
    "risk": "low"
}
```

图 5 蜜罐日志示例

Figure 5 Example of honeypot log

给定特征化表示的安全事件和威胁 IP 样本作为模型输入: $X=(x_1, x_2, \ldots, x_n)$, $x_1, x_2, \ldots, x_n$ 为样本在各维度的特征表示。输出事件或 IP 的威胁程度得分 y。公式表示如下:

$$\widehat{y_{\text{ev}}} = f\left(X_{\text{ev}}\right)$$

其中, X_{ev} 为安全事件的样本表示, y_{ev} 为模型输出的安全事件威胁得分。

$$\widehat{y_{\text{ip}}} = \eta f\left(X_{\text{ip}}\right)$$

其中, X_{ip} 为威胁 IP 的样本表示, y_{ip} 为模型输出的 IP 威胁程度得分, η 为去虚警系数, 由是否为白名单 IP、目的 ip 是否失陷、是否包含攻击 cve 等特征确定。$f(\cdot)$ 为参数已被训练好的回归函数。

基于上述维度设计与特征表示, 给出基于特征增强的威胁评估算法:

算法 1.基于特征增强的威胁评估算法.

输入: 指定周期内所有类型的安全日志集合 D, 需要进行威胁评估的安全事件或威胁 IP 列表 L

输出: 安全事件或威胁 IP 的威胁值

1 根据日志集合 D 中不同类型的日志, 选择威胁评估维度相关字段, 进行字段取值统计

2 参照 4.1 节设定不同字段取值对应的威胁程度分值

3 从日志集合 D 中采样安全事件和威胁 IP, 进行威胁值标注

4 根据标注样本训练回归函数 $f(\cdot)$

5 对标注样本进行 K-means 聚类, 得到聚类中心 c

6 FOR 每个安全事件或威胁 IP 样本 m IN L:

7 对 m 依照 4.1 节的表示规则进行样本表示, 得到 X_m

8 计算 m 与 c 的距离, 增加到 m 的特征表示, 得到 X_m'

9 将 X_m' 输入 $f(\cdot)$, 得到 m 的威胁分值 $f(X_m')$

10 IF m 为威胁 IP:

11 对其进行去虚警, $\eta f(X_m')$作为其威胁分值

12 end FOR

4.4 不同的回归算法

由于威胁程度得分为连续变量, 因此采用回归模型建模上述问题中的函数 $f(\cdot)$。本研究中采用的回归算法有: 线性回归、决策树、随机森林、XGBoost(eXtreme Gradient Boosting)、SVR(Support Vector Regression)、多层感知机。对踩蜜事件、WAF 告警事件和威胁 IP 应用线性回归分析, 此外对 WAF 告警事件应用其他回归算法对比分析。对各回归算法建模分析的详细介绍如下:

(1) 线性回归(Linear Regression, LR)

线性回归模型将各维度特征加权求和得到威胁程度得分。将踩蜜事件、告警事件、威胁 IP 在各维度的表示作为自变量, 将威胁分数作为因变量进行线性回归分析, 公式如下:

$$f\left(x\right) = \sum_{i=1}^{n} \omega_i x_i + b$$

其中, ω_i 为维度的权重系数, b 为偏置项。

(2) 决策树(Decision Tree, DT)

一颗决策树包含一个根节点、若干个内部节点和若干个叶节点。叶节点对应于决策结果, 其他每个节点则对应于一个属性测试; 每个节点包含的样本集合根据属性测试的结果被划分到子节点中; 根节点包含样本全集, 从根节点到每个叶子节点的路径对应了一个判定测试序列。决策树将特征空间划分为 M 个子空间: $R_1, R_2, \ldots, R_M$, C_m 为集合 R_m 中的样本 x_i 对应的输出值 y_i 的均值, N_m 为集合 R_m 中样本的个数。$II(\cdot)$为 0/1 函数。将踩蜜事件、告警事件、威胁 IP 在各维度的表示作为自变量, 将威胁分数作为因变量进行线性回归分析, 公式如下:

$$f\left(x\right) = \sum_{m=1}^{M} C_m II\left(x \in R_m\right)$$

其中, $C_m = \dfrac{1}{N_m} \sum_{i \in N_m} y_i$。

(3) 随机森林(Random Forest, RF)

随机森林是一种集成算法, 它是一个包含 T 个决策树的分类器, 相对于单棵决策树来讲, 随机森林算法通常会有更好的表现, 并能有效防止过拟合现象。随机森林在预测时, 每棵决策树都会产生一个回归结果 $f_i(\cdot)$, 各决策树回归结果的加权平均作为最

终结果。将踩蜜事件、告警事件、威胁 IP 在各维度的表示作为自变量, 将威胁分数作为因变量进行线性回归分析, 公式如下:

$$f(x)=\sum_{i=1}^{T}\omega_i f_i(x)$$

(4) XGBoost(eXtreme Gradient Boosting)

XGBoost 是一种梯度提升算法, 它通过梯度下降迭代训练弱分类器(通常是决策树, 选择最佳分割点来分裂数据), 每一轮迭代都校正上一轮模型的误差, 并集成它们的预测, 逐步提升整体模型性能, 从而构建一个更强大的模型。将踩蜜事件、告警事件、威胁 IP 在各维度的表示作为自变量, 将威胁分数作为因变量进行 XGBoost 建模, 下面公式代表了第 t 次迭代的模型输出:

$$f(x)=\sum_{k=1}^{t}f_k(x)$$

(5) 支持向量回归

支持向量机(Support Vector Machine, SVM)用间隔带而非直线对数据进行拟合, 相比于一般的线性回归, 数据在间隔带内不计算损失, 当且仅当预测值与标签值差距的绝对值大于容忍偏差时才计算损失。SVR 通过最大化间隔带的宽度与最小化总损失来优化模型。将踩蜜事件、告警事件、威胁 IP 在各维度的表示作为自变量, 将威胁分数作为因变量进行支持向量回归建模, 支持向量回归公式如下:

$$f(x)=\sum_{i\in \mathrm{SV}}\left(\alpha_i-\alpha_i^{*}\right)K(x_i,x)+b$$

$$K(x_i,x_j)=\Phi(x_i)^{T}\Phi(x_j)$$

其中, α_i 和 α_i*为拉格朗日系数, $0\le\alpha_i$, $\alpha_i^{*}\le C$(C 为惩罚项), 且 $\alpha_i\alpha_i^{*}=0$。SV 为支持向量集合。$\Phi(\cdot)$为升维函数。

(6) 多层感知机

多层感知机器(Multilayer Perceptron, MLP)是一种神经网络算法, 使用反向传播(Back-Propagation, BP)算法进行训练。基本 BP 算法包括信号的前向传播和误差的反向传播两个过程, 计算误差输出时按从输入到输出的方向进行, 而调整权值和阈值则从输出到输入的方向进行。将踩蜜事件、告警事件、威胁 IP 在各维度的表示作为自变量, 将威胁分数作为因变量进行 MLP 建模, 公式如下:

$$f(x)=\omega_2 g\left(\omega_1^{\mathrm{T}}x+b_1\right)+b_2$$

其中, $g(\cdot)$为激活函数。

4.5 评估效果度量指标

回归分析用到的指标有: p 值[26]、R^2[27]、VIF[28]。p 值用于判断分析项是否出现显著性, 若 p 值小于 0.05 或 0.01, 则说明自变量 X 对因变量 Y 有影响关系; R^2 代表模型对数据的拟合情况, 越大表示拟合效果越好; VIF 用于判断是否存在共线性问题, VIF 值一般>5 说明有共线性问题。线性回归通常使用 F 检验来检验回归模型是否有意义。如果模型通过 F 检验(p<0.05), 说明模型有意义, 至少有一个自变量 X 会对因变量 Y 产生影响; 如果模型没有通过 F 检验(p>0.05), 说明模型构建无意义, X 均不会对 Y 产生影响。回归分析常用的指标介绍如下。

- ***p*:** 判断分析项是否出现显著性, 代表自变量维度对因变量维度存在影响关系, 越小代表影响程度越大。
- **R 方(R-Square):** 决定系数。代表模型拟合程度。

$$R^2=1-\frac{\sum_i\left(\widehat{y_i}-y_i\right)^2}{\sum_i\left(\overline{y_i}-y_i\right)^2}$$

其中, 分子部分表示真实值与预测值的平方差之和; 分母部分表示真实值与均值的平方差之和。根据 R^2 的取值, 来判断模型的好坏, 其取值范围为[0,1]: 如果结果是 0, 说明模型拟合效果很差; 如果结果是 1, 说明模型无错误; 如 0.5 则表示模型有 50%的拟合程度。一般来说, R^2 越大, 表示模型拟合效果越好。R^2 反映的是大概准确率。随着样本数量的增加, R^2 必然增加, 无法真正定量说明准确程度, 只作为定量参考。

- **调整后 R 方(Adjusted R-Square):** 校正决定系数。代表模型拟合程度。

$$\mathrm{R2\_adjusted}=1-\frac{\left(1-R^2\right)(n-1)}{n-p-1}$$

其中, n 是样本数量, p 是特征数量。调整后 R^2 抵消了样本数量对 R^2 的影响, 做到了真正的 0～1, 越大越好。调整 R^2 值用于惩罚任意放置过多的 X, 通常情况下使用较少。若是单变量线性回归, 则使用 R^2 评估, 多变量则使用 R^2_adjusted。

- **均方误差 MSE(Mean Squared Error):** L2 损失, 误差平方和均值, 越接近于 0 越好。

$$\mathrm{MSE}=\frac{1}{m}\sum_{i=1}^{m}\left(y_i-\widehat{y_i}\right)^2$$

- **均方根误差 RMSE(Root Mean Squard Error):** MSE 的平方根。

$$\mathrm{RMSE}=\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i-\widehat{y_i}\right)^2}$$

- **平均绝对误差 MAE(Mean Absolute Error):** L1

损失, 真实值与拟合值的平均值差值, 越接近于 0 越好。

$$\mathrm{MAE}=\frac{1}{m}\sum_{i=1}^{m}\left|\left(y_i-\widehat{y_i}\right)\right|$$

R^2、MSE、RMSE、MAE 根据不同数据会有不同的值大小, 还需结合其他指标和实际情况分析。

4.6 Themis 评估框架的准确性探讨

本研究所提威胁评估框架 Themis 用于对海量告警进行量化评估, 依据量化评估结果对低威胁告警进行过滤, 保留关键告警用于人工研判。

目前在企业的安全运营中, 进行告警过滤通常根据预定义的若干场景优先筛选出特定类型的告警, 即重点关注匹配“高置信度规则”或“高风险规则”的告警。此外, 还可以将多数同类或相关的告警聚合呈现、进行告警分诊等。呈现给安全运营人员的告警推荐则是尽可能地将告警按重要程度排序后的结果[29]。然而, 告警筛选是一个与具体场景有关的、需要大量外部知识的、非常复杂的过程。现阶段可以认为, 不存在任何已公开的、机械的筛选方法, 在实战场景中对安全告警的筛选准确率能够超过经验丰富的人类专家。

本文所提评估框架中的评估维度, 选取现阶段具有代表性的告警数据类型进行设计得到, 各维度下不同的取值对应的威胁分值的设定强依赖于专家人工, 并与选取的回归算法强相关, 无法保证在任何场景下的绝对准确。本研究的创新之处在于, 综合告警日志信息的全部维度从回归分析的角度进行威胁程度的量化, 相较于匹配告警类型等规则式告警过滤方法更具全面性。Themis 框架的准确性主要由两方面决定: 一是严谨的威胁维度与分值设定, 二是良好的回归拟合算法。本研究在实验中对比了现有的常用回归算法的效果差异并给出结论建议; 而对于威胁维度与分值的设计, 则需要借助更为丰富的告警数据与人工专家经验进行迭代优化。

目前在企业实践中, 已经存在一些对告警筛选方法性能和价值的评估指标, 可用于对 Themis 的准确性度量。(1) TopN 精度: 选定某个待评估的时间范围, 让人类专家研判该范围内的、所有被告警筛选方法认定为“关键”的告警, 其中确实属于关键告警的占比即精确率指标。(2) 反馈数据验证: 运营人员对满足某些告警筛选条件的告警进行全量检查, 并反馈研判和处置的结果。反馈数据中的关键告警占比可以认为是衡量告警筛选方法的指标之一。(3) 召回率曲线: 根据运营人员和告警筛选方法发现的关键告警数量绘制召回率曲线, 直观反映告警筛选方法在不同阈值条件下的性能, 以及相比传统安全运营方法的效率提升幅度[22]。

本研究的实验中采用基本的回归指标评估回归模型的效果。此外, 对威胁评估方法的准确性度量可以借鉴上述对告警筛选方法的评估指标, 评估威胁评估方法相比于人工运营的准确度与效率提升。但需要说明的是, 现阶段所有的性能指标均需要人工参与, 因此都无法做到完全客观。因此本文所提威胁量化评估框架, 还需要结合实际场景需求介入人工专家经验进行不断迭代优化。

5 实验与结果

本研究从蜜罐日志和 WAF 日志中采样安全事件样本和威胁 IP 样本, 请安全专家进行威胁值标注, 构建安全事件和威胁 IP 数据集。然后应用回归模型建模, 探究各维度的影响程度, 分析各算法的效果。

5.1 采样与标注

依托研究团队的研究项目, 目前可获取到的同一时间范围的日志有蜜罐日志和 WAF 日志两种, 为项目支撑的某次重保活动中产生的告警日志。蜜罐日志和 WAF 日志均为 Web 层的告警, 可以实现对应用层的入侵检测。在应用层入侵检测中, WAF 是最为广泛使用的传统安全防护设备, 而蜜罐则是主动威胁探测设备的代表。因此选择这两类告警日志进行研究。首先根据告警所属不同的告警类型进行采样, 然后对采样告警进行威胁值的人工标注。

本研究从蜜罐日志和 WAF 日志中采样安全事件样本和威胁 IP 样本。威胁评估对象包括安全事件和威胁 IP。对于安全事件样本采样, 我们从蜜罐和 WAF 的日志中分别抽取 100 个和 200 安全事件样本, 为了使样本覆盖不同的安全事件类型, 我们设定采样的样本分布如表 4 和表 5 所示。保证构建的数据集中包含所有的攻击类型, 避免样本不均衡问题。对于威胁 IP, 从威胁 IP 列表(从日志中构建, 来自踩蜜和 WAF 威胁 IP 列表各 50 个)中抽取 100 个 IP。对于所有样本, 采用[1,10]连续数值作为威胁分值标签, 数值越高代表威胁程度越大。

对告警事件和威胁 IP 进行采样后, 请安全专家进行威胁值的人工标注。对告警事件的标注标准为: 给定某条告警日志, 安全专家对告警日志的信息根据自身经验给出其威胁值。威胁值采用[1,10]连续数值作为威胁分值标签, 数值越高代表该告警事件的威胁程度越大。对威胁 IP 的标注则需要将威胁 IP 在蜜罐日志和 WAF 日志中的所有记录提取出来, 然后安全专家根

表 4 蜜罐日志采样分布

Table 4 Sampling distribution of honeypot logs

踩蜜行为类型	采样数量
普通访问行为	44
触发系统绊线	1
疑似恶意操作	35
网络测绘	10
用户登录行为	8
爬虫	2

表 5 WAF 告警采样分布

Table 5 Sampling distribution WAF alarm logs

WAF 攻击类型	采样数量
HTTP GET 请求访问	4
HTTP POST 请求访问	2
HTTP HEAD 请求访问	1
HTTP OPTIONS 请求访问	1
尝试 SYN 握手	10
IP 黑名单	10
IP 惩罚	10
信息防泄露	20
CC 策略拦截	31
恶意扫描	10
木马后门攻击	10
XSS 攻击	10
SQL 注入攻击	10
命令注入攻击	10
XML 注入攻击	10
服务端模版注入漏洞	10
核心文件非法访问	10
文件上传攻击	10
WEB 应用漏洞攻击	10
未授权访问漏洞	10
其他漏洞防护	1

据该 IP 的日志行为依据自身经验进行威胁值标注, 威胁值范围同样为[1,10]中的连续数值。为了减少人工标注的主观偏差, 将两位安全专家对同一数据的标注结果取均值作为数据标签。告警事件和威胁 IP 标注完成后, 即可应用线性回归进行建模。

对告警事件和威胁 IP 进行标注后, 采用 SPSSAU 在线分析平台[①]进行回归分析。由于本研究是对威胁实体的威胁程度定量评估, 因此实验任务设定为回归任务而非分类任务。将所有标注数据作为训练集输入模型, 模型根据输入样本和威胁值标签迭代优化至稳定状态后, 依据 4.5 节的指标判断模型对各类样本的威胁量化评估拟合效果。

① https://www.spssau.com

5.2 线性回归效果

表 6 展示了线性回归模型对蜜罐事件、WAF 事件和威胁 IP 的威胁程度的建模效果。首先使用 F 检验进行总体显著性检验。从表 6 可知, 多元线性回归通过总体显著性检验($p = 0.000 < 0.05$), 回归模型是有意义的, 说明至少有 1 个自变量维度会对因变量(威胁分数)产生影响, 构建模型有应用价值。

表 6 三种威胁实体的线性回归效果

Table 6 Linear Regression effects of three kinds of threat entities

数据集	p	R^2	调整后 R^2	RMSE
蜜罐事件	0.000	0.772	0.757	0.430
WAF 告警事件	0.000	0.613	0.594	0.764
威胁 IP	0.000	0.564	0.536	1.908

对于蜜罐日志中的安全事件, 将时间、目标域名、目标端口、协议、行为类型、风险等级作为自变量, 将蜜罐事件分数作为因变量进行线性回归分析。从表中可以看出, 模型 R^2 值为 0.772, 意味着六个自变量维度可以解释蜜罐事件分数的 77.2%变化原因。对于 WAF 日志中的安全事件, 将时间、接收时间、代理、目标域名、目标 url、目标端口、攻击描述、位置、运营商作为自变量, 将 WAF 事件分数作为因变量进行线性回归分析。从表中可以看出, 模型 R^2 值为 0.613, 意味着 9 个自变量维度可以解释 WAF 事件分数的 61.3%变化原因。对于威胁 IP, 将出现情况、行为告警数、攻击种类、蜜罐行为、WAF 行为、地理位置作为自变量, 将分数作为因变量进行线性回归分析。从表中可以看出, 模型 R^2 值为 0.564, 意味着六个自变量维度可以解释分数的 56.4%变化原因。

从 R^2、调整后 R^2、RMSE 三项指标来看, 线性回归模型拟合效果最好的是蜜罐事件、其次是 WAF 告警事件、最后是威胁 IP。蜜罐日志中已经给出对安全事件的粗粒威胁量化判断结果, 因此用线性回归拟合相对效果较好。而 WAF 告警依赖的 WAF 设备定义的攻击行为种类较多且无定量威胁评估的初步判断, 因此模型拟合稍差。对于威胁 IP, 评估维度更复杂, 因此线性回归效果较差, 需要增大样本量进一步测试或尝试其他建模方法。

5.3 维度重要性分析

安全事件和威胁 IP 的不同维度对威胁评估的影响程度不同。图 5~图 7 展示了对安全事件和威胁 IP

威胁程度有影响的维度的权重系数, 接近 0 表示影响程度越小。对于蜜罐事件影响不显著的维度有 dst、dport; 对于 WAF 事件影响不显著的维度有 dst、uaAgent; 而威胁 IP 各维度影响较均衡, 无接近 0(即无显著影响的)维度。

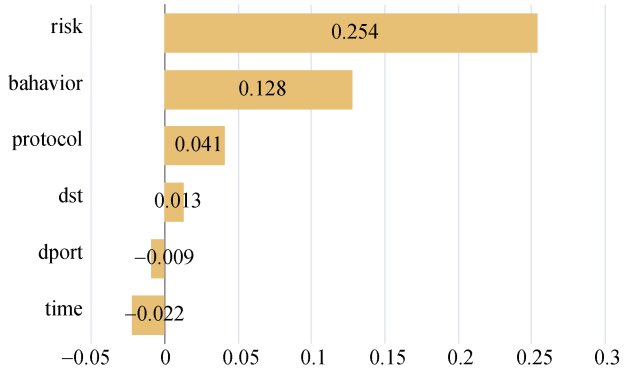


图 5 蜜罐事件维度权重系数柱状图

Figure 5 Bar chart of dimension weight coefficients for honey stepping events

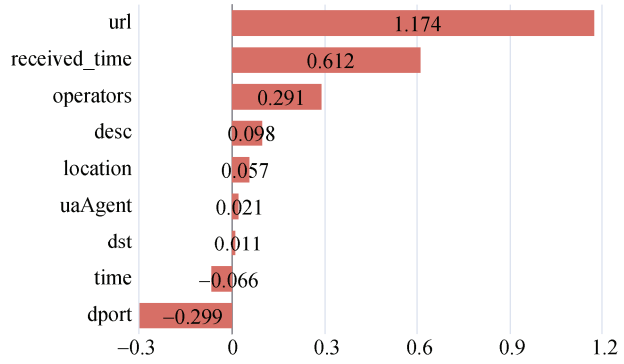


图 6 WAF 告警事件维度权重系数柱状图

Figure 6 Bar chart of dimension weight coefficients for WAF alarm events

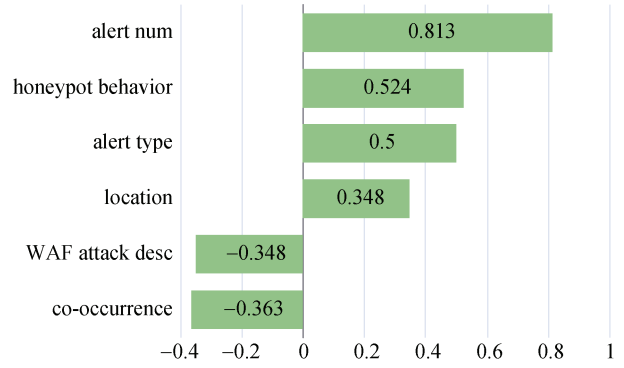


图 7 威胁 IP 的维度权重系数柱状图

Figure 7 Bar chart of dimension weight coefficients for threat IP

在表 7 中, 标记*符号的维度为对威胁评估产生显著影响的维度, 两个*代表该维度与威胁分数具有极高的相关性($p < 0.01$), 一个*代表该维度与威胁分数具有较高的相关性($p < 0.05$)。对蜜罐事件威胁程度影响最大的维度是 risk、behavior, 即蜜罐日志本身自带的威胁程度评级已表示了其威胁程度。对于 WAF 事件威胁程度影响最大的维度是 received_time、time、operators 和 desc, 即威胁程度与攻击时间、运营商、攻击类型强相关。对威胁 IP 威胁程度影响最大的三个维度是 alert_num 和 honeypot behavior, 即引发了较多的报警数、并具有高危踩蜜行为的 IP 威胁程度较高。

表 7 各维度回归分析情况

Table 7 Regression analysis of each dimension

数据集	特征	权重系数	p	VIF
蜜罐事件	time	−0.022	0.386	1.639
	dst	0.013	0.759	1.038
	dport	−0.009	0.939	1.012
	protocol	0.041	0.615	2.389
	behavior	0.128	0.001**	3.525
	risk	0.254	0.000**	4.718
WAF 告警事件	time	−0.066	0.005**	1.081
	received time	0.612	0.000**	2.792
	uaAgent	0.021	0.915	1.025
	dst	0.011	0.823	1.215
	url	1.174	0.145	1.041
	dport	−0.299	0.291	1.025
	desc	0.098	0.011*	2.864
	location	0.057	0.191	1.247
	operators	0.291	0.003**	1.376
威胁 IP	co-occurrence	−0.363	0.412	30.935
	alert num	0.818	0.000**	1.205
	alert type	0.500	0.175	19.206
	honeypot behavior	0.524	0.004**	2.185
	WAF attack desc	−0.348	0.277	10.345
	location	0.348	0.085	1.238

VIF 用于判断共线性问题, 该值小于 5 说明无共线性问题。对蜜罐和 WAF 事件的建模不存在共线性问题。对威胁 IP 的回归建模有三个变量存在共线性问题: co-occurrence、 alert type、 WAF attack desc。威胁 IP 在不同日志中的出现情况和告警数量存在相关性, 因此产生了共线性问题, 该问题可使用岭回归或者逐步回归进行解决。对于 WAF attack desc 的共线性还需进一步探究。模型构建完成后, 可依据 4.3 节的公式进行对安全事件或 IP 的威胁评估。

5.4 特征增强消融实验

表 8 展示了在无特征增强与带有特征增强的情况下, 模型在威胁 IP 回归分析中的结果对比。从整体看, 加入特征增强后, 模型的 R^2 与调整后 R^2 均有所提升, 证明使用增强后的特征表示有助于提升模型在威胁 IP 数据整体的拟合能力。从各维度看, 进行特征增强后, 其中三个维度的 p 值有下降, 证明增强后的特征可以提

升这些维度的显著性。

表 8 威胁 IP 数据集上的特征增强消融实验

Table 8 Ablation experiment of feature enhancement on threat IPs

特征和指标	p(无特征增强)	p(带有特征增强)
co-occurrence	0.412	0.409↓
alert number	0.000**	0.000**
alert type number	0.175	0.111↓
honeypot behavior	0.004**	0.002**↓
WAF attack desc	0.277	0.289
location	0.085	0.259
R^2	0.564	0.572↑
调整后 R^2	0.536	0.539↑

5.5 不同回归算法比较

除了线性回归, 我们在 WAF 告警日志数据上对比了不同的回归算法在各个指标上的效果。对算法的介绍见 4.3 节。对 R^2、MAE、MSE、RMSE 四个指标接介绍见 4.4 节, 对 MAD、MAPE、EVS、MSLE 的指标介绍如下。

- 中位数绝对误差 MAD(Median Absolute Deviation): 预测值离中位数残差的绝对值, 不受异常值影响, 越小越好。
- 平均绝对百分误差 MAPE(Mean Absolute Percentage Error): 平均误差百分比, 不受异常点影响, 越小越好。
- 可解释方差分 EVS(Explained Variance Score): 衡量模型对数据波动的解释力度, 介于[0,1]之间, 越大越好。
- 均方根对数误差 MSLE(Mean Squared Logarithmic Error): 在 RMSE 相同时, 其对欠预测惩罚更多。

图 8 和图 9 两个柱状图分别展示了基于各机器学习回归算法: LR(线性回归)、DT(决策树)、RF(随机森林)、XGBoost、SVR(支持向量回归)、MLP(多层感知机)的威胁评估模型在不同指标上的结果。图 8 显示了各个模型在 R^2、MAE、MSE、RMSE 指标上的表现; 图 9 显示了各个模型在 MAD、MAPE、EVS 和 MSLE 上的表现。横坐标的指标箭头代表该指标的调优方向, 箭头向上表示模型在该指标的数值越大越好, 箭头向下代表模型在该指标的数值越小越好。

首先, 可以看到在 R-squared 指标上, LR 的表现最好, DT、RF、XGBoost 以及 SVR 的效果相当, 效果最差的是 MLP, R-squared 值仅为 16.5%, 建模意义不大。其次, 在 MAE, MSE 和 RMSE 指标上, 各模型表现也与在 R^2 的比较结论相同: DT、RF、XGBoost 以及 SVR 的效果相当, MLP 效果最差, 模型的预测误差最大。最后, 在 MAD, MAPE, EVS 和 MSLE 指标上, MLP 仍然显著逊色于其他模型, DT、RF、XGBoost 以及 SVR 相差无几。从上述现象可以得出, 在数据量不大、数据结构不复杂的情况下, 可优先选用线性回归建模。在数据量较小的场景下, DT、RF、SVR 以及 XGBoost 的效果差异不大, 可优先选用复杂度低的算法。MLP 作为一种神经网络算法, 训练复杂度高, 在数据量较少的情况下拟合效果不好。然而, 受限于标注样本过少, 本实验尚未能评估在数据量较大情况下各回归算法的效果, 需在后续研究中进行重点分析。

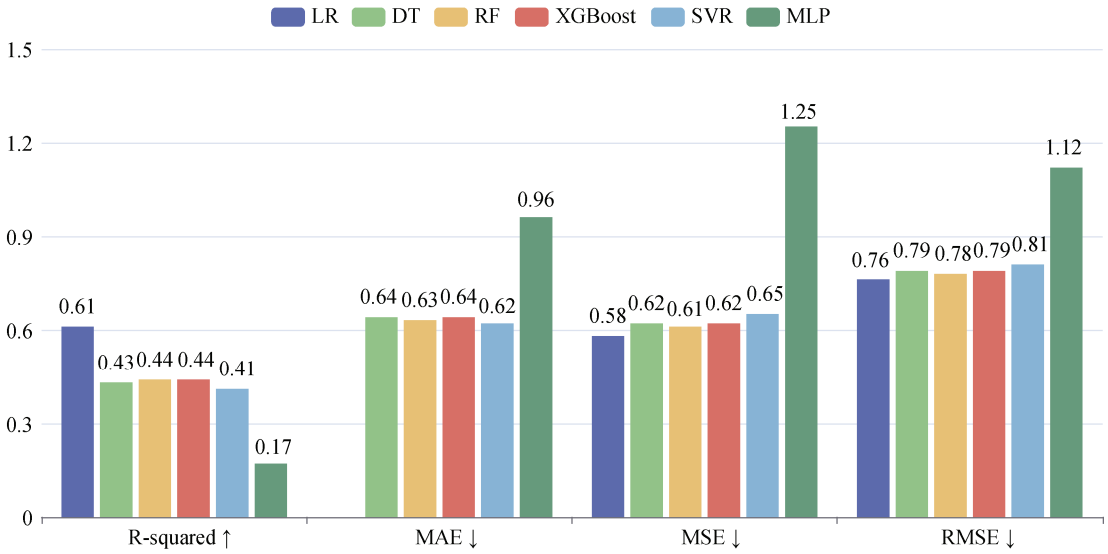


图 8 各回归算法在各指标上的效果比较(1)

Figure 8 Performance comparison of the regression algorithms on each metric (1)

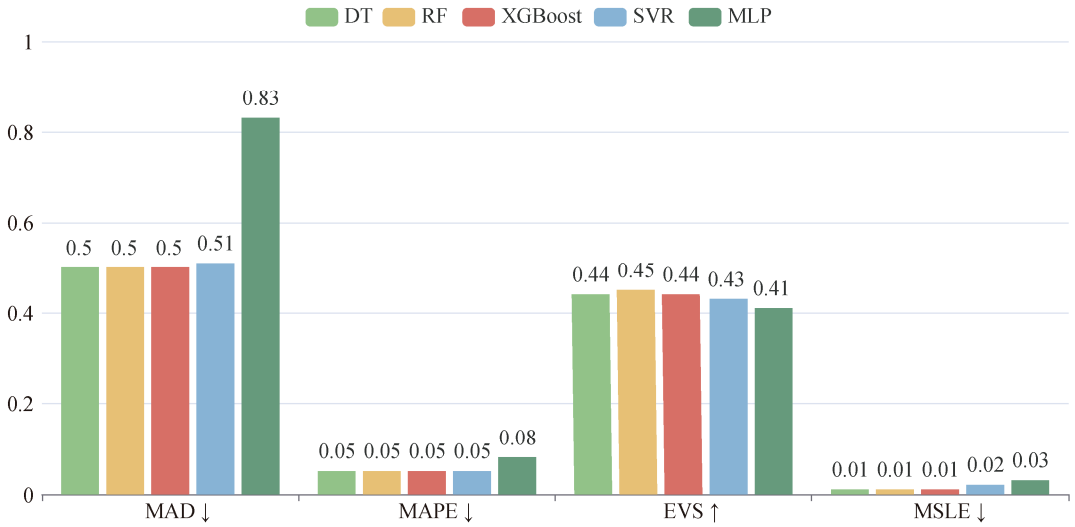


图 9　各回归算法在各指标上的效果比较(2)

Figure 9　Performance comparison of the regression algorithms on each metric (2)

5.6　复杂度分析与应用建议

表 9 展示了各机器学习算法在训练阶段的时间复杂度。其中, N 代表样本数量, m 代表特征维度。对于线性回归, 复杂度与样本数 N 和特征数 m 的平方成正比。决策树的复杂度与特征数 m、样本数 N 的平方以及样本数的对数 $\log N$ 成正比。随机森林由多个决策树组成, 其中 M 代表树的数量, N 代表样本数量。与单个决策树相比, 随机森林中每个树是并行构建的, 复杂度与树的数量、样本数和样本数的对数成正比。XGBoost 复杂度与随机森林相似, 与树的数量 M、样本数 N 和样本数的对数 $\log N$ 成正比。SVR 的复杂度与特征数 m 和样本数 N 的平方成正比。MLP 是一种神经网络, 其复杂度与多个因素有关, 包括样本数 N、输入层特征数 m、隐藏层数 k、每个隐藏层的神经元数 h、输出层神经元数 o 以及迭代次数 i。由于涉及多个参数, MLP 的复杂度可能会非常高, 特别是在网络结构复杂或数据量大时。

表 9　各回归算法的训练复杂度

Table 9　Training complexity of regression algorithms

机器学习回归算法	训练复杂度
线性回归	$O(N \cdot m^2)$
决策树	$O(m \cdot N \cdot \log N)$
随机森林	$O(M \cdot m \cdot N \cdot \log N)$
XGBoost	$O(M \cdot m \cdot N \cdot \log N)$
SVR	$O(m \cdot N^2)$
MLP	$O(N \cdot m \cdot h^k \cdot o \cdot i)$

上述分析可以看出, 各个算法训练复杂度的影响因素各不相同, 有的算法虽然时间复杂度低, 但却以空间复杂度的提高作为代价。根据 5.5 节中各种算法的效果比较, 在样本量和特征数均较少时, 建议优先选用线性回归建模。SVR 复杂度比决策树更低, 但效果无明显下降, 可以考虑采用。鉴于决策树与集成算法随机森林和 XGBoost 的效果相差无几, 而集成算法需要更高的空间复杂度, 建议在几种算法中优先选用决策树。而神经网络算法的复杂度受多种因素影响而效果较差, 建议在样本量较小时暂不选用。

基于本研究的发现, 对相关安全运维团队和安全研究人员给出如下建议:

(1) 在告警数量较多、安全分析资源有限时, 重点关注对威胁程度评估存在显著性影响的威胁维度(蜜罐日志的踩蜜行为描述、WAF 告警类型描述、威胁 IP 的告警数量等)下的信息。

(2) 结合实际资源与场景需求选择合适的回归算法(线性回归 > 支持向量回归 > 决策树 > 集成算法(随机森林、XGBoost) > 多层感知机), 以平衡模型的效果和资源性能。

6　结论与讨论

针对 Web 层的攻击类型多种多样, 基于规则和分类算法的入侵检测系统应对海量攻击告警存在局限。本研究提出 Themis 威胁评估框架, 首创性地采用回归方法对 Web 安全日志中的威胁进行量化评估, 通过定量分析细粒度评估威胁实体的威胁程度。我们基于多来源多类型的安全日志, 对不同的告警事件和威胁 IP 定义了相应的威胁维度, 并对威胁样本进行特征表示和无监督下的特征增强, 构建回归模

型计算威胁实体的威胁程度得分。实验对各维度信息在威胁评估的显著性影响进行了分析, 得出安全运维中应当侧重关注的威胁维度, 同时验证了无监督特征增强方法在威胁评估中的效果。此外, 对常规的机器学习回归算法进行了效果测试和复杂度分析, 在小样本不均衡数据场景下, 推荐优先采用线性回归进行威胁评估建模。

然而, 在实际应用过程中, 我们仍面临一些挑战, 如标签准确性的提升、样本量的增加以及评估模型的进一步优化等, 需要在如下方面进行改进。(1) 采样完整性: 确保采样过程及样本量能够全面覆盖各种类型的安全事件和威胁 IP, 避免样本偏差, 增强模型的泛化能力。(2) 评分科学性: 建立更加科学的评分体系, 引入更多的安全专家进行交叉评分, 确保评分结果能够真实反映威胁实体的实际威胁程度。优化威胁评分系统, 提供更细致的分值粒度, 以更精确地反映威胁程度。(3) 定制化模型: 根据不同应用场景下能获取到的日志类型和特点, 定制化威胁评估模型, 以更好地适应特定的应用环境。

参考文献

[1] Hassan W U, Guo S J, Li D, et al. NoDoze: Combatting Threat Alert Fatigue with Automated Provenance Triage[C]. *Proceedings 2019 Network and Distributed System Security Symposium*, 2019. DOI:10.14722/ndss.2019.23349.

[2] Hossain M N, Milajerdi S M, Wang J N, et al. SLEUTH: Real-Time Attack Scenario Reconstruction from COTS Audit Data[EB/OL]. 2018: arXiv: 1801.02062. https://arxiv.org/abs/1801.02062.

[3] Milajerdi S M, Eshete B, Gjomemo R, et al. POIROT: Aligning Attack Behavior with Kernel Audit Records for Cyber Threat Hunting[C]. *The 2019 ACM SIGSAC Conference on Computer and Communications Security*, 2019: 1795-1812.

[4] Xiao R Z, Chen H, Lu J T, et al. AllInfoLog: Robust Diverse Anomalies Detection Based on all Log Features[J]. *IEEE Transactions on Network and Service Management*, 2023, 20(3): 2529-2543.

[5] Sun Y Z, Guo S M, Chen Z W. Intelligent Log Analysis System for Massive and Multi-Source Security Logs: MMSLAS Design and Implementation Plan[C]. *2019 15th International Conference on Mobile Ad-Hoc and Sensor Networks*, 2019: 416-421.

[6] Xu J C, Shu X K, Li Z. Understanding and Bridging the Gap between Unsupervised Network Representation Learning and Security Analytics[C]. *2024 IEEE Symposium on Security and Privacy*, 2024: 3590-3608.

[7] He J J, Tang C, Li W S, et al. BR-HIDF: An Anti-Sparsity and Effective Host Intrusion Detection Framework Based on Multi-Granularity Feature Extraction[J]. *IEEE Transactions on Information Forensics and Security*, 2024, 19: 485-499.

[8] Wang Xiaoyu, Gong Xiaorui, Zhang Xiu, et al. A Survey of Multi-step Attack Detection[J]. *Journal of Cyber Security,* 2023.
(王笑语, 龚晓锐, 章秀, 等. 多步攻击检测技术综述[J]. *信息安全学报*, 2023.)

[9] Han Xueying, Wang Zehui, Liu Runshi, et al. Overview of Advanced Persistent Threats Detection Technology[J]. *Journal of Cyber Security,* 2024.
(韩雪莹, 王泽辉, 刘润时, 等. 高级持续性威胁检测技术研究综述[J]. *信息安全学报*, 2024.)

[10] Zeng Q W, Zhang G M, Xing C Y, et al. Intelligent Attack Path Discovery Based on Heuristic Reward Shaping Method[J]. *Journal of Cyber Security*, 2024, 9(3): 44-58.
(曾庆伟, 张国敏, 邢长友, 等. 基于启发式奖赏塑形方法的智能化攻击路径发现[J]. *信息安全学报*, 2024, 9(3): 44-58.)

[11] Katipally R, Yang L, Liu A Y. Attacker Behavior Analysis in Multi-Stage Attack Detection System[C]. *The Seventh Annual Workshop on Cyber Security and Information Intelligence Research*, 2011: 1.

[12] Jia Y, Gu Z Q, Du L, et al. Artificial Intelligence Enabled Cyber Security Defense for Smart Cities: A Novel Attack Detection Framework Based on the MDATA Model[J]. *Knowledge-Based Systems*, 2023, 276: 110781.

[13] Milajerdi S M, Gjomemo R, Eshete B, et al. HOLMES: Real-Time APT Detection through Correlation of Suspicious Information Flows[C]. *2019 IEEE Symposium on Security and Privacy*, 2019: 1137-1152.

[14] Xiong C L, Zhu T T, Dong W H, et al. Conan: A Practical Real-Time APT Detection System with High Accuracy and Efficiency[J]. *IEEE Transactions on Dependable and Secure Computing*, 2022, 19(1): 551-565.

[15] Li Y F, Gao Y, Ayoade G, et al. Heterogeneous Domain Adaptation for Multistream Classification on Cyber Threat Data[J]. *IEEE Transactions on Dependable and Secure Computing*, 2024, 21(1): 1-11.

[16] Yu Y Q, Yan H B, Ma Y, et al. DeepHTTP: Anomalous HTTP Traffic Detection and Malicious Pattern Mining Based on Deep Learning[C]. *Cyber Security*, 2020: 141-161.

[17] Zang X D, Gong J, Wang M L, et al. IP Traffic Behavior Characterization *via* Semantic Mining[J]. *Journal of Network and Computer Applications*, 2023, 213: 103603.

[18] Yan H N, Li X G, Zhang W J, et al. Automatic Evasion of Machine Learning-Based Network Intrusion Detection Systems[J]. *IEEE Transactions on Dependable and Secure Computing*, 2024, 21(1): 153-167.

[19] Diallo A F, Patras P. Adaptive Clustering-Based Malicious Traffic Classification at the Network Edge[C]. *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications*, 2021: 1-10.

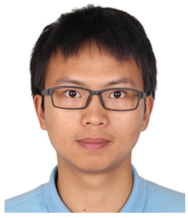
[20] Wei N, Yin L H, Zhou X M, et al. A Feature Enhancement-Based Model for the Malicious Traffic Detection with Small-Scale Imbalanced Dataset[J]. *Information Sciences*, 2023, 647: 119512.

[21] Ding H W, Sun Y, Huang N N, et al. TMG-GAN: Generative Adversarial Networks-Based Imbalanced Learning for Network Intrusion Detection[J]. *IEEE Transactions on Information Forensics and Security*, 2024, 19: 1156-1167.

[22] Yang D Q, Liu W M, Yu Z. Research on Active Defense Application Based on sHoneypot[J]. *Chinese Journal of Network and In-*

formation Security, 2018, 4(1): 57-62.
(杨德全, 刘卫民, 俞宙. 基于蜜罐的主动防御应用研究[J]. *网络与信息安全学报*, 2018, 4(1): 57-62.)
[23] 蜜罐和蜜网. https://c4pr1c3.github.io/cuc-ns/chap0x11/main.html. Oct. 2023.
[24] Koenker R, Hallock K F. Quantile Regression[J]. *The Journal of Economic Perspectives*, 2001, 15(4): 143-156.
[25] Cade B S, Noon B R. A Gentle Introduction to Quantile Regression for Ecologists[J]. *Frontiers in Ecology and the Environment*, 2003, 1(8): 412.
[26] Fisher R A. Statistical Methods for Research Workers[M]. Breakthroughs in Statistics, 1992.
[27] Student. The Probable Error of a Mean[M]. Breakthroughs in Statistics, 1992.
[28] Johnston J. Econometric methods 3rd ed[M]. McGraw-Hill Book Company, 1984.
[29] AISecOps: 量化评估告警筛选方案的性能. NSFOCUS. https://blog.nsfocus.net/aisecops/. Oct. 2022.

冯文英 于 2023 年在中国科学院大学网络空间安全专业获得博士学位。现于鹏城实验室新型网络研究部任助理研究员。研究领域为安全事件检测与分析、知识图谱。研究兴趣包括: 安全日志分析、多步攻击检测、威胁情报与攻击组织归因等。Email: fengwy@pcl.ac.cn



顾钊铨 于2015年在清华大学计算机科学与技术专业获得博士学位。现任哈尔滨工业大学(深圳)教授, 鹏城实验室双聘研究员, 鹏城网络靶场总师。研究领域为网络靶场、安全事件检测与分析。CCF 高级会员。研究兴趣包括: 网络安全态势感知、人工智能安全。Email: guzhaoquan@hit.edu.cn

赵昂霄 于 2024 年在电子科技大学计算机技术专业获得硕士学位。现于哈尔滨工业大学(深圳)计算机科学与技术专业攻读博士学位。研究领域为网络空间安全。研究兴趣包括: 网络靶场、知识图谱和人工智能安全。Email: zhaoax@pcl.ac.cn

罗翠 于 2009 年在西安电子科技大学计算机科学与技术专业获得硕士学位。现任鹏城实验室高级工程师。研究领域为网络空间安全。研究兴趣包括: 网络安全数据分析、攻击检测。Email: luoc@pcl.ac.cn

袁华平 于 2019 年在广东工业大学计算机科学与技术专业获得硕士学位。现任鹏城实验室新型网络研究部助理工程师。研究领域为网络安全。研究兴趣包括: 机器学习、知识图谱。Email: yuanhp@pcl.ac.cn

胡宁 博士, 博士生导师, 鹏城实验室新型网络研究部研究员, 毕业于国防科技大学计算机科学与技术专业。CCF 高级会员。主要研究方向包括: 网络空间安全、工业控制系统安全、软件定义网络与网络虚拟化等。Email: hun@pcl.ac.cn

【附录】

附表 1 蜜罐日志字段介绍

Appendix 1 Introduction to honeypot log fields

蜜罐日志字段	字段描述	蜜罐日志字段	字段描述
_index	该条日志的索引名称, 即唯一标识	sport	源端口号, 源设备使用的端口
_type	该条日志的文档类型	dport	目标端口号, 目标设备开放的服务端口
_id	文档唯一标识符, 与“_index”相同	sub_device_type	子设备类型, 为“蜜罐”
_score	日志搜索结果的相关性得分	action	动作类型, 如 TCP 连接尝试
_source	原始文档数据区域	payload	数据传输的有效负载, 此处为空对象
_class	类名, 表示记录所属的类	attck	攻击描述, 如"尝试 SYN 握手"等描述。
time	事件发生的时间戳, 单位毫秒	env	环境标签, 例如"beta", 表示测试或预发布环境
src	源 IP 地址, 发起请求或连接的设备 IP	desc	事件描述, 如"普通访问行为"
dst	目标 IP 地址或域名, 被访问或连接的目标设备或服务	behavior	行为描述, 同样表示为"普通访问行为"
protocol	通信协议类型, 如 TCP、UDP 等	threat_level	威胁级别, 数值型, 数值大小代表威胁程度
risk	风险评估, 分为高中低三级, 如“low”, 表示风险较低	honeyAction	对于蜜罐系统的操作描述, 同样表示为"普通访问行为"

附表 2　WAF 日志字段介绍

Appendix 2　Introduction to WAF log fields

WAF 日志字段	字段描述	WAF 日志字段	字段描述
_index	该条日志的索引名称, 即唯一标识	src	源 IP 地址, 发起请求的客户端 IP
_type	该条日志的文档类型	dst	目标 IP 地址或域名, 被访问的服务端地址, 与 domain 相同
_id	文档唯一标识符, 与“_index”相同	sport	源端口号
_score	日志搜索结果的相关性得分	dport	目标端口号
_source	原始文档数据区域	logType	日志类型, 这里是“Syslog”, 表示系统日志
_class	类名, 表示记录所属的类	device	设备类型, 为“WAF”, 即 Web 应用防火墙
uaAgent	用户代理字符串, 即访问客户端的浏览器信息	received_time	日志接收时间戳, 单位为毫秒, 表示服务器接收到日志的时间
domain	访问的域名	desc	事件描述, 如“CC 策略拦截”, 表示触发了 WAF 的 CC 攻击防护策略
loginStatus	登录状态, 数值型	location	客户端地理位置
cve	可能存在的 CVE 漏洞列表	longitude	经度坐标, 表示客户端所在地理位置的经度
url	访问的完整 URL, 与 domain 相同	latitude	纬度坐标, 表示客户端所在地理位置的纬度
time	事件发生的时间戳, 单位为毫秒	operators	运营商信息, 表示客户端使用的互联网服务提供商