基于多模态融合与动态对抗训练的智能网络安全检测系统研究

罗 彩[1] 王 勇[1]
LUO Cai WANG Yong

摘 要 针对传统网络入侵检测模型在复杂场景下的检测精度较低、鲁棒性不足等问题，文章提出了一种多模态融合 + 动态对抗检测系统，系统采用轻量化的四层架构，依托改进 VDBERT 垂域模型实现流量时序与日志语义特征深度融合，搭配 GAN 对抗训练提升模型抗变异攻击鲁棒性。实验选取 CIC-IDS2017 与 CERNET-Log 混合数据集开展多组测试，并进一步对 DDoS、SQL 注入、端口扫描等典型攻击类别的分类结果进行对比分析。实验结果表明，多模态融合动态对抗检测系统在混合数据集上的准确率、精确率、召回率、F_1-score 等指标均有不同程度的提升，误报率得到有效降低，综合性能优于各类基准模型，能够有效满足公开与真实网络场景的入侵检测需求，为复杂网络环境下的安全防护提供技术支撑。

关键词 人工智能；网络安全检测；多模态融合；动态对抗训练；生成对抗网络

doi：10.3969/j.issn.1672-9528.2026.07.022

0 引言

随着互联网技术的快速普及，网络已成为社会生产生活的核心基础设施，与此同时，DDoS 攻击、SQL 注入、端口扫描等各类网络攻击行为频发，严重威胁网络安全、数据隐私与系统稳定 [1]。网络入侵检测作为网络安全防护体系的核心环节，其核心任务是实时识别网络中的异常行为，及时预警攻击风险，为网络安全防护提供决策依据。

目前，网络入侵检测模型主要分为传统机器学习模型与深度学习模型两大类。传统机器学习模型如随机森林（RF）、支持向量机（SVM）等，优点是训练速度快、可解释性强，但在处理高维网络流量、复杂攻击场景时，容易出现特征提取不充分、检测精度较低的问题。深度学习模型，比如 CNN、VDBERT 等，能通过深层网络结构实现特征的自动提取，在入侵检测中展现出较好的性能，但单一模态的深度学习模型很难同时融合网络流量、系统日志等多源数据的特征，面对新型隐蔽攻击时鲁棒性不足。

针对上述问题，本文提出了一种多模态融合 + 动态对抗检测系统，整合网络流量与系统日志多源数据特征，通过动态对抗训练提升模型对新型攻击、隐蔽攻击的识别能力。为了验证模型的有效性与优越性，本文采用公开标准数据集与真实场景数据相结合的实验方案，以 CIC-IDS2017 数据集为基准，以 CERNET-Log 教育网真实日志数据集为真实场景验证，并与主流检测模型展开对照实验，全面评估所提模型的检测性能，为网络入侵检测技术的优化与应用提供一些实验依据与理论参考。

1. 广西国际商务职业技术学院 广西南宁 530007

[基金项目] 广西国际商务职业技术学院 2024 年度校级科研项目“基于机器学习的网络安全攻击检测系统研究”（GSB2024020）

[6] 杨航 , 樊凯 , 梁段 . 基于蚁群算法的电力数据网络 APT 攻击特征分析及防御技术 [J]. 微型电脑应用 , 2023, 39(7): 101-104.

[7] 孙勇 , 刘高原 . 针对电网审计业务应用的网络攻击检测 [J]. 微型电脑应用 ,2024,40(3):89-92.

[8] 高国庆 , 徐天竹 , 屈巍 , 等 . 电力信息系统网络安全态势智能评估方法 [J]. 微型电脑应用 ,2024,40(11):145-148.

[9] 吴莉艳 , 孙开元 , 陈坤 , 等 . 基于 CNN-LSSVM 的电力系统虚假数据攻击检测 [J]. 浙江电力 ,2024,43(11):90-96.

[10] 钟睿 , 冯文英 , 李若南 , 等 . 基于 ATT&CK 框架的技术关联分析方法综述 [J]. 网络空间安全科学学报 , 2025, 3(5): 2-13.

【作者简介】

童加斌（1986—），男，福建长汀人，硕士，讲师，研究方向：计算机技术、教育技术。

（收稿日期：2026-06-17 修回日期：2026-07-16）

1 相关技术与研究现状

1.1 核心技术基础

多模态融合技术通过整合异构数据源的特征信息，弥补单一模态数据缺陷，按照融合阶段可分为数据级、特征级、决策级三类。其中特征级融合兼顾信息完整性与计算效率，通过提取各模态核心特征再加权融合，适配网络安全检测的实时性需求。跨模态注意力机制可自主筛选关键特征，提升融合精度，是本文特征融合环节的核心技术。

生成对抗网络（GAN）由生成器与判别器构成，通过二者博弈训练生成仿真攻击样本，可扩充训练数据集、优化模型泛化能力，能够提升检测系统对抗变异攻击的能力。VDBERT 是网络安全垂域预训练模型，其中注入了 CVE 漏洞库、ATT&CK 攻击战术等领域知识，可以更加高效地提取日志语义与流量时序特征。

1.2 相关工作

近年来，国内外学者围绕网络入侵检测模型展开了大量研究。在传统机器学习领域，冷斌在随机森林构建防御系统基础上，对网络安全事件进行实时监测、自动化响应和持续学习，提高了网络安全防御的效率和准确性 [1]。

在深度学习领域，CNN 模型凭借局部特征提取优势，被广泛应用于网络流量检测。Al-zubidi 等 [2] 提出一种混合深度学习模型，实现 DoS 与 DDoS 攻击检测。张慧妍 [3] 则提出了一种基于深度强化学习的网络异常流量检测模型，解决了当前网络环境中攻击样本标记困难导致无法检测未知攻击或模型过拟合的问题，提高了入侵检测能力。

2 系统架构设计

本文提出的“多模态融合＋动态对抗检测系统”采用“数据预处理—多模态融合—动态检测—自适应响应”四层轻量化架构。系统核心是通过多模态特征挖掘弥补单一数据缺陷，借助动态对抗训练强化模型鲁棒性，实现威胁精准识别与快速处置，整体架构如图 1 所示。

2.1 模块功能设计

2.1.1 数据预处理模块

负责采集网络流量、系统日志两类核心数据，完成清洗、归一化与特征格式化。流量数据提取包大小、传输频率、连接时长等 18 维时序特征，采用 *Z*-score 标准化映射至 [0,1] 区间；日志数据经分词、去停用词处理后，转换为语义特征向量，剔除冗余噪声、填补缺失值，为后续融合环节筑牢数据基础。

2.1.2 多模态融合模块

以改进 VDBERT 垂域模型为核心，拆分单模态特征提取与跨模态注意力融合两步。采用 CNN 提取流量时空特征，Transformer 架构提取日志语义特征，通过余弦相似度计算注意力权重，实现两类特征的深度加权融合，用公式表示为：

$$\alpha_i = \frac{\exp\left(\mathrm{sim}(\boldsymbol{f}_{\mathrm{flow},i}, \boldsymbol{f}_{\mathrm{log},i})\right)}{\sum_{j=1}^{n} \exp\left(\mathrm{sim}(\boldsymbol{f}_{\mathrm{flow},j}, \boldsymbol{f}_{\mathrm{log},j})\right)} \tag{1}$$

式中：$\boldsymbol{f}_{\mathrm{flow},i}$ 为流量第 i 维特征，$\boldsymbol{f}_{\mathrm{log},i}$ 为日志第 i 维特征，$\mathrm{sim}(\cdot)$ 为余弦相似度函数，α_i 为第 i 个特征对应的注意力权重。通过该机制突出关键特征权重，弱化冗余信息干扰，提升特征表达有效性。

从作用机理看，注意力机制的核心价值在于对不同模态、不同维度特征的重要性进行动态评估。网络攻击行为往往同时表现为流量侧的异常波动与日志侧的语义异常，例如 DDoS 攻击更容易在连接频率、包大小分布、会话持续时间等时序特征中形成显著模式，而 SQL 注入则更多体现在 URL 参数、数据库报错、访问路径和日志语义中。若采用简单拼接或平均融合，容易使关键特征被大量正常或弱相关特征稀释，导致模型对隐蔽攻击和低频攻击的识别能力下降。

加权融合则在注意力权重的基础上对流量特征和日志语义特征进行自适应组合，使高相关、高区分度特征在最终融合向量中占据更高权重，低贡献或噪声特征被适当抑制。由此，模型既能够保留流量时序特征对大规模异常行为的敏感性，又能够利用日志语义特征对攻击意图和攻击路径的解释能力，从而提升分类边界的清晰度与跨场景泛化能力。该设计也是本文模型在混合数据集和真实场景数据中保持较高召回率、较低误报率的重要原因。

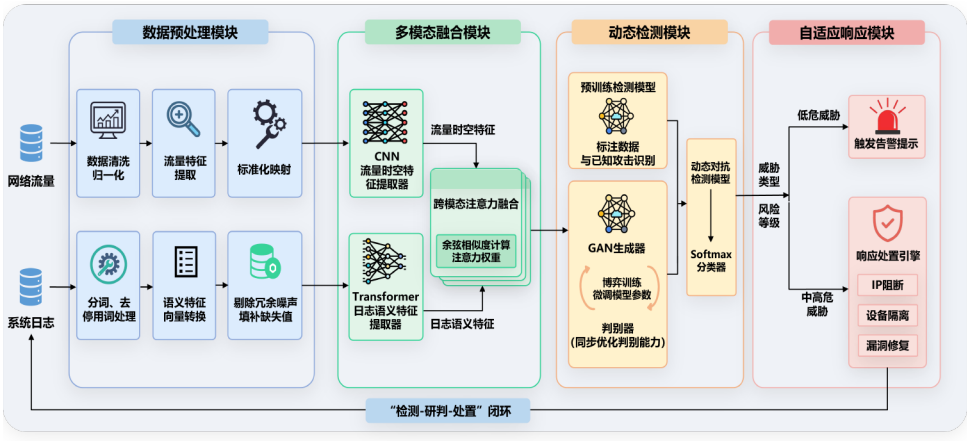


图 1 模型整体架构

2.1.3 动态检测模块

搭载预训练检测模型与 GAN 对抗训练单元，预训练模型基于标注数据实现已知攻击快速识别；GAN 生成器模拟变异攻击样本，判别器同步优化判别能力，通过博弈训练微调模型参数，提升对零日攻击、变异威胁的识别能力，最终通过 Softmax 分类器输出威胁类型与风险等级。

2.1.4 自适应响应模块

依据检测结果进行分级处置，低危威胁会触发告警提示，中高危威胁则会执行 IP 阻断、设备隔离、漏洞修复等动作，形成“检测—研判—处置”链路，缩短攻击响应周期。

2.2 系统工作流程

系统运行遵循“数据采集—预处理—特征融合—动态检测—响应处置”的流程。首先采集流量与日志原始数据，经预处理生成标准化特征；接着通过多模态模块完成特征深度融合；随后送入检测引擎完成威胁识别与等级判定；最后触发对应响应策略，同时将结果上传至管理界面。

3 实验过程与结果分析

3.1 实验环境与数据集

基于 Windows 11 操作系统，CPU 采用的 Intel(R) Core(TM) i7-10710U CPU @ 1.10 GHz；内存为 16 GB。编程环境：Python3.8.20，所用框架：PyTorch2.4.0。

本文采用公开标准数据集与真实场景数据结合的方案，CIC-IDS2017 数据集涵盖 DDoS、SQL 注入等 14 类常规攻击，共 280 万条样本；CERNET-Log 数据集包含真实日志与新型攻击样本，共 50 万条。两个原始数据集分别按 7:2:1 比例划分训练集、验证集和测试集；在综合测试阶段，为兼顾常规攻击与未知 / 变异攻击覆盖度，从两类测试数据中抽取并组合 36 万条样本构成混合测试集（正常样本 20 万条、攻击样本 16 万条）。数据集详情如表 1 所示。

表 1 实验数据集详情

数据集名称	样本总量 / 万条	正常样本 / 万条	攻击样本 / 万条	攻击类型
CIC-IDS2017	280	160	120	14 类常规攻击
CERNET-Log	50	32	18	新型变异攻击
混合测试集	36	20	16	常规 + 未知攻击

3.2 评价指标

选取准确率（Accuracy）、精确率（Precision）、召回率（Recall）、F_1-score、误报率（false positive rate, FPR）五项核心指标，全面评估系统性能。

准确率（Accuracy）定义为：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{2}$$

精确率（Precision）定义为：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{3}$$

召回率（Recall）定义为：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{4}$$

F_1-score 定义为：

$$F_1\text{-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

误报率（FPR）定义为：

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \times 100\% \tag{6}$$

式中：TP 为真阳性，TN 为真阴性，FP 为假阳性，FN 为假阴性，指标数值直观反映系统检测精度、误报控制与实时性水平。

3.3 对比实验与结果分析

3.3.1 CIC-IDS2017 数据集实验结果

CIC-IDS2017 数据集作为网络入侵检测的公开基准数据集，用于验证各模型的基础检测性能。CIC-IDS2017 数据集，包含多种攻击类型，如 DoS/DDoS、Brute Force、Heartbleed 等，以及正常流量数据。实验结果如表 2 所示。

表 2 CIC-IDS2017 数据集实验结果

单位：%

检测模型	准确率	精确率	召回率	F_1-score	误报率
随机森林（RF）	95.2	94.8	94.5	94.6	4.8
SVM（RBF 核）	89.4	88.7	88.1	88.4	10.6
1D-CNN	96.5	93.2	90.1	91.6	5.4
VDBERT	96.1	95.8	95.5	95.6	3.9
多模态融合 + 动态对抗检测	96.7	96.5	96.3	96.4	3.3

由表 2 可以看出，在 CIC-IDS2017 数据集上，多模态融合 + 动态对抗检测系统的各项性能指标均最优，准确率为 96.7%，F_1-score 为 96.4%，误报率为 3.3%，优于其他四种基准模型。其中，VDBERT 模型的性能仅次于目标模型，其 F_1-score 为 95.6%，主要得益于其强大的语义特征提取能力。SVM 模型的性能最差，准确率仅 89.4%，误报率为 10.6%，主要由于其在高维流量数据处理中易出现过拟合，泛化能力不足。可以发现，基于深度学习的检测模型，其检测性能要普遍优于传统机器学习的模型。

3.3.2 CERNET-Log 数据集实验结果

CERNET-Log 数据集包含真实日志与新型攻击样本，用

于验证各模型在真实场景下的鲁棒性与实际适配性，实验结果如表 3 所示。

表 3 CERNET-Log 数据集实验结果

单位：%

检测模型	准确率	精确率	召回率	F_1-score	误报率
随机森林（RF）	92.3	91.8	91.5	91.6	10.6
SVM（RBF 核）	84.7	83.9	83.2	83.5	5.4
1D-CNN	91.4	89.7	88.9	89.3	3.9
VDBERT	93.8	93.5	93.1	93.3	1.3
多模态融合 + 动态对抗检测	94.9	94.6	93.9	94.4	2.1

由表 3 的结果可以看出，在真实场景的 CERNET-Log 数据集上，1D-CNN 的召回率仅 88.9%，对新型攻击的识别能力不足。多模态融合 + 动态对抗检测系统在准确率、精确率、召回率和 F_1-score 四项指标上均为最高，其中准确率为 94.9%，F_1-score 为 94.4%；其误报率为 2.1%，虽高于 VDBERT 的 1.3%，但低于随机森林、SVM 和 CNN 模型。该结果表明，目标系统在真实场景中以略高于 VDBERT 的误报率换取了更强的攻击识别能力与综合 F_1 表现，整体仍具有较好的场景适配性。与 CIC-IDS2017 数据集的实验结果相比，所有模型的性能均有一定下降，主要由于 CERNET-Log 数据集中包含新型隐蔽攻击样本，且日志数据的复杂性高于公开数据集。

3.3.3 混合数据集实验结果

混合数据集由 CIC-IDS2017 与 CERNET-Log 数据集合并而成，用于综合验证各模型在基准与真实场景下的综合性能，实验结果如表 4 所示。

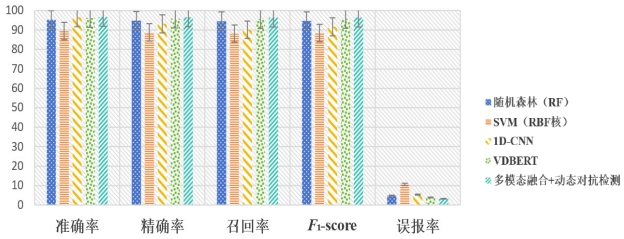
表 4 混合数据集实验结果

单位：%

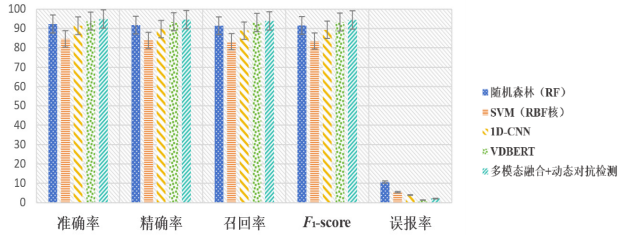
检测模型	准确率	精确率	召回率	F_1-score	误报率
随机森林（RF）	93.8	93.4	93.1	93.2	6.2
SVM（RBF 核）	87.1	86.5	86.0	86.2	12.9
1D-CNN	93.2	92.8	92.4	92.6	6.8
VDBERT	95.0	94.7	94.3	94.5	5.0
多模态融合 + 动态对抗检测	95.6	95.1	94.8	95.0	3.7

由表 4 可知，在混合数据集上，多模态融合 + 动态对抗检测系统的综合性能依然最优，准确率为 95.6%，F_1-score 为 95.0%，误报率为 3.7%，既保持了在公开基准数据集上的高精度，又具备在真实场景中的高鲁棒性。

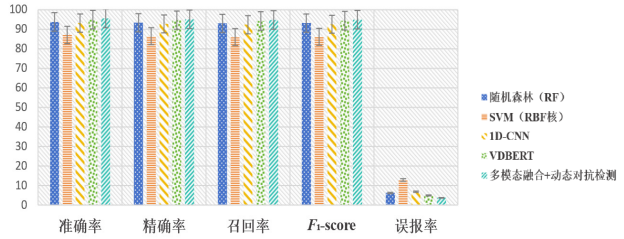
三个数据集的实验数据对比结果如图 2 所示。



（a）CIC-IDS2017 数据集实验数据对比



（b）CERNET-Log 数据集实验数据对比



（c）混合数据集实验数据对比

图 2 实验数据对比

3.3.4 典型攻击类型分类结果对比

为进一步回应不同攻击类型检测效果的差异性问题，本文在混合测试集中选取 DDoS、SQL 注入、端口扫描、暴力破解及新型变异攻击等具有代表性的攻击样本，统计多模态融合 + 动态对抗检测系统在不同攻击类别上的分类表现，结果如表 5 所示。

表 5 典型攻击类型分类结果对比

单位：%

攻击类型	样本来源	精确率	召回率	F_1-score
DDoS	CIC-IDS2017	97.3	96.8	97.0
SQL 注入	CIC-IDS2017/ CERNET-Log	94.8	93.6	94.2
端口扫描	CIC-IDS2017	95.9	95.1	95.5
暴力破解	CIC-IDS2017	94.6	93.9	94.2
新型变异攻击	CERNET-Log	92.7	91.8	92.2

由表 5 可知，系统对 DDoS、端口扫描等流量模式明显的攻击类型具有较高识别能力，其中 DDoS 攻击的 F_1-score 达到 97.0%，端口扫描的 F_1-score 达到 95.5%，说明流量时序特征能够较好地刻画此类攻击的行为规律。对于 SQL 注入和暴力破解等更依赖日志语义和行为上下文的攻击类型，系

统仍保持较高的精确率和召回率，表明日志语义特征与流量特征的融合能够有效提升对复杂攻击意图的捕捉能力。

同时，新型变异攻击的 F_1-score 为 92.2%，低于常规攻击类别，但相较单一模态模型仍具有更好的识别稳定性。这说明动态对抗训练能够在一定程度上增强模型对未知攻击和变异攻击的适应能力，但在攻击样本稀缺、攻击行为高度隐蔽的情况下，模型仍需要结合更多真实场景样本持续迭代优化。几种典型攻击类型检测结果对比如图 3 所示。

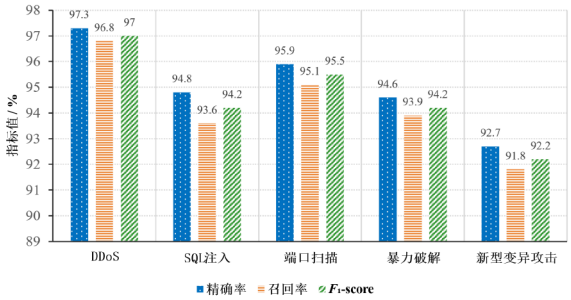


图 3 几种典型攻击类型检测结果对比

3.4 实验结论

实验数据表明，多模态融合 + 动态对抗检测系统在公开基准数据集、真实场景数据集及混合数据集上整体表现较优。在 CIC-IDS2017 和混合数据集上，目标系统各项核心指标均优于基准模型；在 CERNET-Log 真实场景数据集上，目标系统在准确率、精确率、召回率和 F_1-score 方面表现最优，但误报率略高于 VDBERT。总体来看，该系统能够在保持较高检测精度和召回能力的同时兼顾较好的误报控制，具备较强的鲁棒性与实用性。

深度学习模型（VDBERT、CNN）的整体性能优于传统机器学习模型（随机森林、SVM），其中 VDBERT 模型的语义特征提取能力突出，在真实场景中表现较好，CNN 模型则在局部特征提取上具有优势。多模态融合技术能够有效整合多源数据的特征优势，动态对抗训练能够提升模型对新型攻击、隐蔽攻击的识别能力，两者结合是提升网络入侵检测性能的有效途径。

从攻击类型维度看，本文模型在 DDoS、端口扫描等流量特征明显的攻击类别上表现更优，在 SQL 注入、暴力破解等需要结合日志语义判断的攻击类别上也保持了较好的检测效果。该结果说明，多模态融合不仅提升了整体检测指标，也增强了模型对不同攻击机理的适配能力，使系统能够在多源异构数据场景下实现更细粒度的威胁识别。

4 结论与展望

针对传统网络入侵检测模型在复杂场景下检测精度不足、鲁棒性不足等问题，本文构建了一种多模态融合 + 动态对抗检测系统，通过改进 VDBERT 模型实现异构特征深度融合，借助 GAN 对抗训练优化模型鲁棒性，提升了模型的检测精度与鲁棒性。以 CIC-IDS2017 公开数据集与 CERNET-Log 真实日志数据集为实验基础，与随机森林、SVM、CNN、VDBERT 等主流模型展开了对照实验。实验结果表明，多模态融合 + 动态对抗检测系统综合性能优于各类基准模型，能够有效识别常规攻击与新型隐蔽攻击，尤其在 DDoS、SQL 注入、端口扫描等典型攻击类别上表现出较好的分类效果，满足公开基准与真实网络场景的入侵检测需求。

从实际部署角度看，系统仍可能面临以下挑战：一是真实网络环境中的流量规模大、业务类型复杂，对模型推理时延、吞吐能力和边缘侧部署效率提出了更高要求；二是不同单位的日志格式、字段规范和采集粒度差异较大，可能影响多模态特征对齐与融合效果；三是攻击者会持续改变攻击策略，导致样本分布漂移和模型性能衰减，需要建立持续学习与定期再训练机制；四是系统在联动阻断、隔离和告警等响应环节还需考虑误报带来的业务影响，避免自动化响应过度干预正常业务。

未来的研究内容将聚焦于以下几个方面：一是增加更多场景的实验数据集，如工业控制网络、物联网等特殊场景数据，验证模型在不同领域的适配性；二是构建轻量化模型结构，降低模型的训练成本，提升系统在实际网络环境中的部署能力。

参考文献：

[1] 冷斌 . 基于人工智能技术的计算机网络安全防御系统设计与实现 [J]. 信息记录材料 ,2024,25(11):91-92.

[2] Al-zubidi A F, Farhan A K, Towfek S M.Predicting DoS and DDoS attacks in network security scenarios using a hybrid deep learning model[J]. Journal of Intelligent Systems, 2024, 33(1): 20230195.

[3] 张慧妍 . 基于记忆模块和过滤式生成对抗网络的入侵检测算法研究 [D]. 山西 : 山西大学 ,2024.

【作者简介】

罗彩（1990—），女，湖南湘乡人，本科，讲师，研究方向：信息安全、深度学习。

王勇（1980—），男，湖北荆州人，本科，工程师，研究方向：信息安全。

（收稿日期：2026-06-12 修回日期：2026-07-15）