基于超参数优化和 LightGBM 算法的 DDoS 攻击检测与分类

胡宏伟，孙皓月
（河北建筑工程学院，河北　张家口　075000）

摘　要：针对分布式拒绝服务攻击（DDoS）数据流量样本容量大、数据特征多的特点以及检测分类准确率低的问题，提出了一种基于 LightGBM（Light Gradient Boosting Machine）算法的 DDoS 攻击检测与分类方法。在 CICDDoS2019 数据集预处理和特征筛选的基础上，构建 LightGBM 检测模型和多分类模型。同时在模型预训练时，采用随机网格搜索与贝叶斯超参数优化技术实现超参数自动调优。实验结果表明，该模型在检测与分类任务上能达到 98.34% 的准确率。该研究为 DDoS 攻击提供了一种高效且简易的检测与分类思路。

关键词：DDoS 攻击；超参数优化；LightBGM；检测与分类
中图分类号：TP393　　**文献标识码：**A　　**DOI：**10.19358/j.issn.2097-1788.2025.07.003
引用格式：胡宏伟，孙皓月. 基于超参数优化和 LightGBM 算法的 DDoS 攻击检测与分类［J］. 网络安全与数据治理，2025，44(7)：15-19，26.

DDoS attack detection and classification based on hyperparameter optimization and LightGBM algorithm

Hu Hongwei，Sun Haoyue
（Hebei University of Architecture，Zhangjiakou 075000，China）

Abstract：Aiming at the characteristics of large sample capacity and multiple data features of distributed denial of service attack（DDoS）data traffic as well as the problem of low detection and classification accuracy，this paper proposes a DDoS attack detection and classification method based on LightGBM（Light Gradient Boosting Machine）algorithm. Based on the preprocessing and feature screening of the CICDDoS2019 dataset，the LightGBM detection model and multi-classification model are constructed. Meanwhile，random grid search and Bayesian hyperparameter optimisation techniques are used to achieve hyperparameters auto-tuning during model pre-training. The experimental results show that the model in this paper can achieve an accuracy rate of 98.34% in the detection and classification tasks. This research aims to provide an efficient and simple detection and classification idea for DDoS attacks.

Key words：DDoS attacks；hyperparameter optimization；LightBGM；detection and classification

0　引言

分布式拒绝服务攻击（DDoS）是一种普遍常见的攻击方式。随着网络信息的传输速度进一步加快以及物联网的发展，DDoS 攻击变得比以往更加活跃，攻击范围及规模也日益扩大[1]。由于 DDoS 攻击的危害性大，每次发生重大攻击事件波及范围广，因此，DDoS 攻击检测始终是网络空间安全领域一个非常重要的研究方向。传统的 DDoS 攻击检测方法通常基于规则和阈值来识别异常流量，但这些方法难以应对新型和复杂的攻击。因此，研究人员一直在寻求更先进更高效的技术来应对这一威胁。机器学习技术已经显示出在网络安全领域中具有巨大潜力，可以识别并分类 DDoS 攻击流量。在此背景下，本研究旨在探索一种基于 LightGBM 算法的 DDoS 攻击检测与分类方法。

1　研究现状

近年来，网络安全问题日益凸显，尤其是 DDoS 攻击对网络造成的威胁逐渐增大。基于快快网络最新发布的 2025 年 DDoS 攻击趋势白皮书显示，2024 年中国全年遭受 DDoS 攻击超 307 万次，同比增长 89.7%，攻击流量峰值达到创纪录的 2.35 Tb/s[2]。更值得警惕的是，AI 技术的深度应用正在重塑攻击模式，使攻击变得更加智能化和自适应。为有效防范 DDoS 攻击，研究人员进行了一系

列关键检测技术的研究。例如王翊铭提出了一种基于统计特征提取和集成学习算法的 DDoS 攻击检测模型[3]，并设计了基于网络数据流量的 DDoS 攻击检测系统，但由于训练集中 DDoS 攻击种类少，该研究没有针对于新类型的 DDoS 攻击方式的检测。蓝瑞童提出了基于自编码器与分布式梯度增强库（XGBoost）相结合的混合检测模型[4]，但该模型需要大量的计算资源和时间进行训练和调整，这可能使得模型在实际应用中不适用快速防护出现的安全漏洞，尤其是对于实时的 DDoS 攻击检测。刘译夫在基于 CNN 的 DDoS 攻击多分类方法的研究与应用中[5]，通过对 SYN Flood 和 ICMP Flood 两种经典的 DDoS 攻击方式进行研究，提出了基于卷积神经网络的 DDoS 多分类方法，但该研究数据类别样本单一，没有对 DDoS 攻击手段进行跟进与研究。针对上述研究存在的不足，本文采用 LightGBM 算法框架，构建 LightGBM 检测模型和多分类模型，该模型可有效地处理高维特征的 DDoS 数据，并且结合超参数优化技术，可以达到较高的检测与分类性能。

2　相关技术

2.1　超参数优化方法

2.1.1　基于随机网格搜索的超参数优化

目前来说 sklearn 中超参数优化器有四种，分别是网格搜索、随机网格搜索、对半网格搜索和对半随机网格搜索[6]。上述四种网格搜索方法适用于超参数空间维度低且每个参数的候选值范围非常有限的场景，其原理是在定义的网格点上保证遍历所有或部分组合并评估每种组合的性能，具有天然高度并行化和评估任务相互独立的特点。同时该方法忽略了参数间的潜在相关性，无法利用评估历史信息智能引导搜索方向。

2.1.2　基于贝叶斯优化器的超参数优化

贝叶斯优化方法[7]是当前超参数优化领域的前沿范式，是当前先进的优化框架。该方法可以被应用于自动化机器学习的各大领域，不止限于超参数搜索的领域，更是可以被用于神经网络架构搜索以及元学习等先进的领域。现代几乎所有在效率和效果上取得优异成果的超参数优化方法都是基于贝叶斯优化的基本理念而形成的，因此贝叶斯优化是整个自动化机器学习中学习的重点。同时，贝叶斯优化是解决中高维参数空间优化和目标函数评估成本高昂问题的首选方法，其适合于计算资源相对有限、但需要在合理迭代次数内找到高性能模型配置的场景。

本文采用随机网格搜索方法与贝叶斯优化方法作对比，验证两种方法对本文模型性能的提升效果。

2.2　LightGBM 算法

LightGBM 是微软推出的一种新的 boosting 框架[8]，其通过损失函数的泰勒展开式近似表达残差，另外利用正则化项来控制模型的复杂度。LightGBM 最大的特点是使用了基于直方图的决策树算法[9]和基于梯度提升[10]的策略，该策略降低了内存的使用，并加速了训练过程。同时通过只选择分裂增益最大的节点进行分裂，避免了某些节点增益较小带来的开销。

2.2.1　单边梯度抽样算法

单边梯度抽样算法（Gradient One-Side Sampling，GOSS）旨在减少计算量，特别是在处理具有大量样本的数据集时，其通过减少每次迭代中用于训练模型的样本数量，来平衡计算效率与模型精度。GOSS 算法保留了梯度较大的样本，对梯度较小的样本进行下采样。同时为了弥补下采样导致的样本分布问题，在下采样的样本上加入放大系数。其在计算信息增益时，可以以少量的样本来代替所有低梯度的样本。

2.2.2　互斥特征绑定算法

互斥特征绑定算法（Exclusive Feature Bundling，EFB）用于减少特征数量，特别是在特征维度很高的情况下，通过捆绑互斥的特征来降低模型的复杂度。EFB 识别出那些在数据集中很少同时取非零值的特征对，并将它们捆绑在一起，这样在决策树的某个节点上只需要一个分裂就可以同时考虑这些特征。EFB 通过减少特征数量，从而减小了模型的复杂度和内存占用，同时加速了训练过程。

3　数据预处理

3.1　数据来源

本实验所选用的公共数据集为加拿大网络安全研究所（CIC）DDoS 数据集——CICDDoS2019，其包含良性和最新的常见 DDoS 攻击，类似于真实世界数据[11]。如表 1 所示，实验选取了七种 DDoS 类型，分别是 UDP、DNS、NTP、SSDP、NetBIOS、LDAP 和 MSSQL。

表 1　实验数据量表

流量类型	样本数量	攻击类别
Benign	50 000	正常网络流量
UDP	50 000	UDP 洪泛
DNS	50 000	DNS 查询洪泛
NTP	50 000	NTP 反射攻击
SSDP	50 000	SSDP 反射攻击
NetBIOS	50 000	NetBIOS 协议滥用
LDAP	50 000	LDAP 反射攻击
MSSQL	50 000	MSSQL 服务洪泛

3.2　DDoS 数据特征选取

在 LightGBM 中，特征重要性对于理解模型对特征的

依赖程度至关重要。原数据中共有 61 种特征，根据特征重要度的结果，筛选重要性较高的特征作为后续模型的训练输入，如图 1 所示。

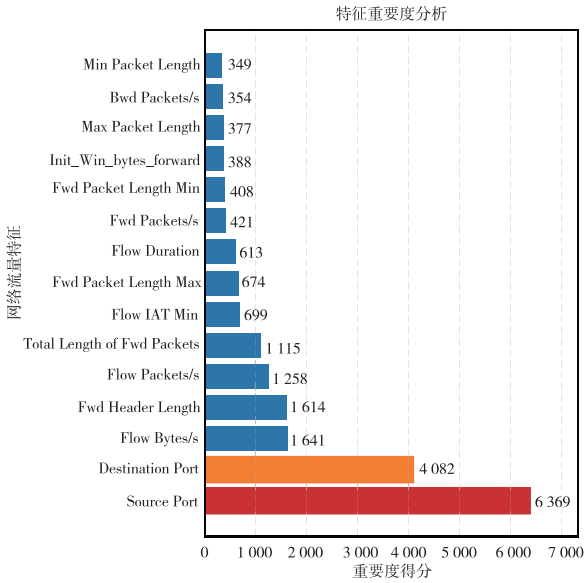

图 1　特征选取图

4　基于 LightGBM 算法的 DDoS 检测与分类

4.1　DDoS 检测与分类流程

在 DDoS 检测任务上，首先使用 LightGBM 框架定义了一个检测模型。这个模型的目标是通过学习训练数据集中的网络流量模式和规律，有效地识别正常流量和异常行为。在训练过程中，利用训练数据集对模型进行拟合，确保模型能够准确地分类正常流量和 DDoS 攻击相关的异常流量，模型能够捕捉到 DDoS 攻击的特征和模式。

模型训练完成后，使用测试数据集来评估模型的性能。在这一阶段，模型将测试数据集中的流量判定为正常或异常，并标记异常流量数据，作为后续 DDoS 类别分类任务的基础。

最后在 DDoS 类别分类任务中，使用相同的 LightGBM 框架，定义并训练了一个多分类模型。该模型旨在区分不同类型的 DDoS 攻击，通过学习训练数据集中的网络流量模式，识别出不同的攻击类别。为了评估多分类模型的性能，使用得到的异常流量作为测试数据集。通过将这些异常流量数据输入到模型中，评估模型在区分 DDoS 攻击类别上的准确性和效果。多分类模型有助于网络管理员更精确地识别和应对不同类型的 DDoS 攻击，提高网络的整体安全性和抗攻击能力。

如图 2 所示，检测与分类任务流程从数据收集和准备开始，包括获取并预处理网络流量数据，随后通过数据探索和特征工程阶段，着重于提取关键特征以供模型训练。接下来，将数据划分为训练集、验证集和测试集，并通过 LightGBM 进行模型训练，通过反复的验证和调整过程提高模型性能。最后，在测试集上评估最终模型，并考虑性能优化措施，以便将模型顺利部署到实际环境中。

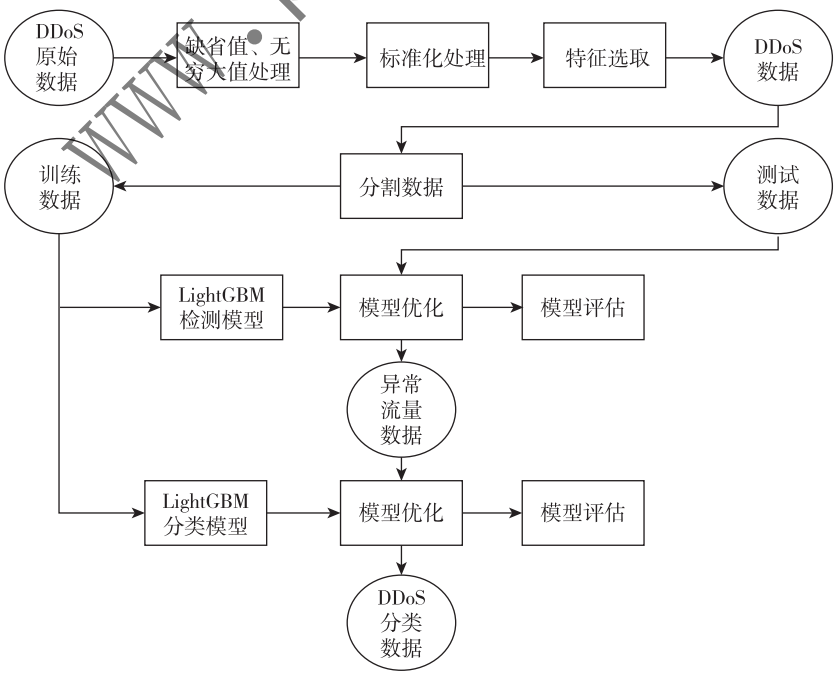

图 2　DDoS 检测与分类流程图

4.2 DDoS 检测与分类实验

为验证本文提出的基于 LightGBM 算法的 DDoS 检测与分类性能，采用多个不同的算法来进行对比实验。对比算法包括 XGBoost、支持向量机（SVM）、随机森林（Random Forest）、线性判别分析（LDA）、K 最近邻（KNN）、高斯朴素贝叶斯（GaussianNB）、深度神经网络（DNN）、决策树（Decision Tree）和自适应增强（AdaBoost）。以上算法均在相同的 DDoS 攻击数据集上进行了训练和测试。

基于 DDoS 数据的不同算法准确率对比如图 3 所示。实验结果表明，LightGBM 算法在对比实验中，准确率达到 92.20%，表明其在处理这种复杂的分类任务时表现尤为出色。这归因于该模型在处理不平衡数据和捕捉非线性关系方面的能力，其在正则化和灵活性方面的优势，使其能提供较好的泛化性能和较高的准确率。

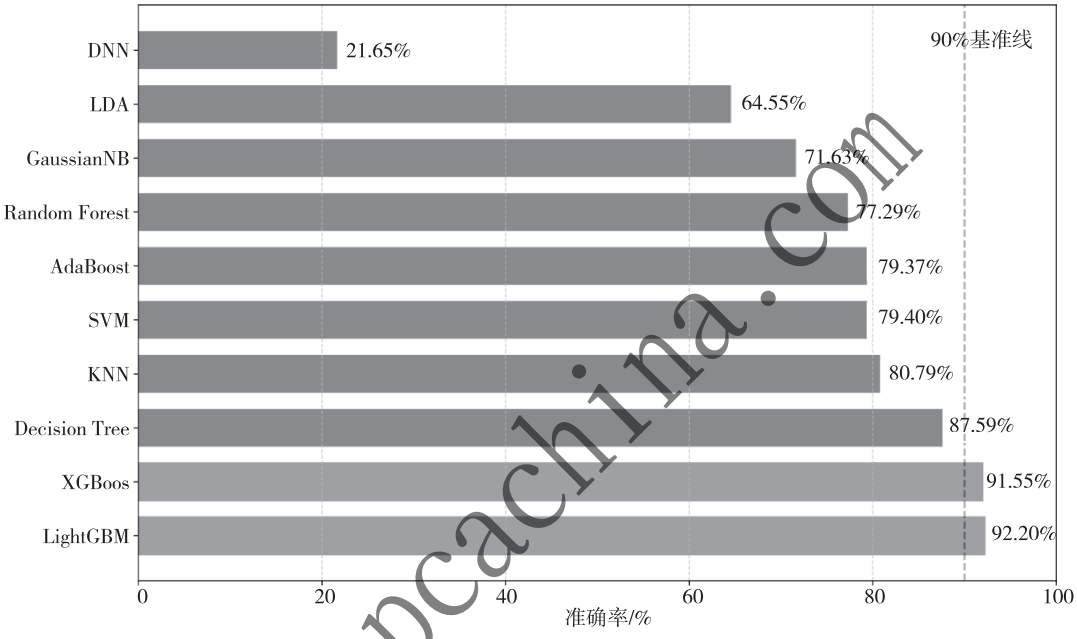

图 3　不同算法准确率对比图

5　基于超参数优化的 LightGBM 模型

LightGBM 作为一种高效的梯度提升决策树算法，凭借其出色的性能在上述检测与分类任务实验中取得最优准确率。然而要充分发挥 LightGBM 的潜力，超参数的选择与优化至关重要。图 4 所示为 LightGBM 模型的超参数优化流程，通过精细的参数调整，旨在进一步提升模型的预测精度与泛化能力。

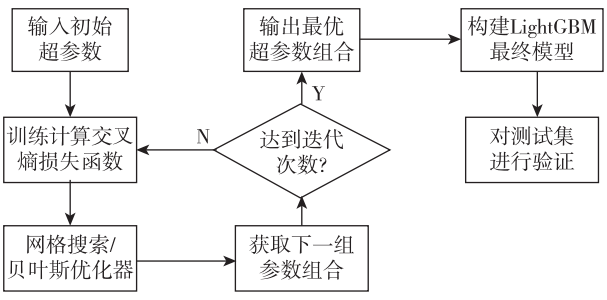

图 4　超参数优化流程

5.1 基于随机网格搜索的超参数优化

对于 LightGBM 检测与分类模型，随机网格搜索可以系统地遍历多个超参数组合，通过交叉验证来评估每个组合的性能，从而找到最优的超参数设置，如学习率、树的深度、叶子节点的数量等[12]。将数据集划分为训练集、验证集和测试集。训练集用于训练模型，验证集用于评估模型的性能并选择最优超参数，测试集用于最终评估模型的泛化能力。使用 Python 的 scikit-learn 库中的 RandomizedSearchCV 类来设置随机网格搜索。设置要优化的超参数及其取值范围，以及交叉验证的折数。调用 RandomizedSearchCV 的 fit 方法，在训练集上训练模型，并在验证集上评估模型性能。RandomizedSearchCV 会随机遍历部分超参数组合，找到最优的组合。根据随机网格搜索的结果，选择验证集上性能最优的超参数组合。使用最优超参数组合在训练集上重新训练模型，得到最终的 LightGBM 检测与分类模型。在测试集上评估最终模型

的性能，以验证其泛化能力。

5.2 基于贝叶斯优化器的超参数优化

选择一个合适的贝叶斯优化库，首先定义需要估计的目标函数及其定义域，然后通过有限观测值对函数进行估计，构建代理函数来近似目标函数，并持续迭代更新估计值。接着定义采集函数来指导下一个采样点的选择，以最大化信息增益。最后，在达到停止条件时，从已观测点中找到最优超参数配置[13]。重复这一过程，直到达到预设的迭代次数或收敛条件。使用该方法找到最优参数组合预训练后的 LightGBM 模型。在测试集上评估最终模型的性能，以验证其泛化能力。

5.3 实验结果与模型评估

本实验采用准确率、精确度、召回率、*F*1 分数作为评价模型性能的指标[14]：

（1）准确率表示分类正确的样本数占总样本数的比例。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \times 100\% \tag{1}$$

（2）精确度为预测为正的样本中实际为正的比例，衡量模型“预测精度”。

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\% \tag{2}$$

（3）召回率为实际为正的样本中被正确预测的比例，衡量模型“覆盖能力”。

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \times 100\% \tag{3}$$

（4）*F*1 分数为精确度和召回率的调和平均数，平衡两者关系。

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \tag{4}$$

在本文不同算法准确率对比实验中的 LightGBM 预训练模型的基础上，采用不同超参数优化的模型性能评估如表 2 所示。LightGBM 在 DDoS 检测与分类任务中取得了 98% 以上的准确率，这标志着其在处理网络流量数据方面表现出色。在参数优化的过程中，虽然随机网格搜索的搜索范围广但是计算成本较高，而贝叶斯优化通过对参数空间进行概率建模，能够更智能地选择下一个尝试的参数组合，从而在相对较少的尝试中找到最优解。实验结果表明，通过参数优化成功地提高了 LightGBM 检测与分类模型的准确率，同时贝叶斯优化相对于传统的随机网格搜索在 LightGBM 的参数调整中表现更为优越，达到了 98.34% 的准确率，进一步提升了 LightGBM 在 DDoS 检测与分类任务中的性能。

表 2 不同超参数优化后 LightGBM 模型性能评估表（%）

	准确率	精确度	召回率	*F*1 分数
原模型（检测）	96.74	94.46	95.31	94.63
随机网格搜索（检测）	96.85	94.65	95.28	94.87
贝叶斯优化器（检测）	97.28	95.05	95.03	95.04
原模型（分类）	97.15	97.37	97.01	97.24
随机网格搜索（分类）	97.56	97.67	97.92	97.76
贝叶斯优化器（分类）	98.34	98.43	98.67	98.55

6 结束语

本文在现有的 DDoS 攻击数据的基础上进行数据处理与特征选取，同时在基于 LightGBM 模型基础上构建检测模型与分类模型，通过设计多个模型对比实验，验证了本文模型在 DDoS 攻击数据检测任务上的优势。最后结合随机网格搜索与贝叶斯优化器的超参数优化方法，提高了 DDoS 攻击的检测与分类任务的性能。通过准确率、精确度、召回率、*F*1 分数进行模型评估，实验结果表明在 DDoS 攻击检测与分类上采用贝叶斯优化器后的 LightGBM 模型较随机网格搜索方法有更高的性能，对于 DDoS 攻击分类准确率达到 98.34%、精确度达到 98.43%、召回率达到 98.67%、*F*1 分数达到 98.55%。未来在本文模型的基础上，可采用 SDN 架构开发混合轻量级防御系统[15]，采用模型蒸馏达到轻量化，并将优化后的检测模型嵌入 SDN 交换机，实现 DDoS 攻击的有效防御。

参考文献

[1] 王博，万良，叶金贤，等. 融合稀疏注意力机制在 DDoS 攻击检测中的应用 [J]. 计算机工程与设计，2024，45（5）：1312－1320.

[2] 快快网络发布《2025 年 DDoS 全球攻击趋势专项报告》[J]. 中国信息安全，2025（5）：98.

[3] 王翊铭. 基于网络流量的 DDoS 攻击检测技术的研究与实现 [D]. 南京：东南大学，2022.

[4] 蓝瑞童. 基于 AE-XGBoost 模型的 DDoS 攻击检测方法研究 [D]. 成都：西南财经大学，2022.

[5] 刘译夫. 基于 CNN 的 DDoS 攻击多分类方法的研究与应用 [D]. 海口：海南大学，2020.

[6] 万飞. 基于网格搜索的支持向量机在入侵检测中的应用 [D]. 合肥：合肥工业大学，2016.

[7] 魏守鑫. 基于改进贝叶斯优化的超参数优化方法的研究与实现 [D]. 西安：西安电子科技大学，2022.

[8] KE G，MENG Q，FINLEY T，et al. LightGBM：a highly efficient gradient boosting decision tree [C] //Advances in Neural Information Processing Systems，2017：3146－3154. （下转第 26 页）

和 Gas-Based 方法。实验结果表明，该方法可有效抑制 MEV 提取行为，显著提高低 Gas 交易的执行机会，并在不显著延长确认时间的前提下降低整体交易成本，展现出良好的实用价值与推广前景。

尽管该方法在模拟环境下取得了较好效果，但仍存在两个值得进一步研究的方向：一是如何提升模型在主网复杂交易环境下的适应能力；二是如何平衡模型训练时延与链上实时性需求。未来研究可考虑引入图神经网络、联邦学习等机制，进一步提升模型泛化能力和部署效率，以应对更加动态和安全敏感的区块链交易场景。

参考文献

[1] BUTERIN V. Ethereum whitepaper [EB/OL]. (2025 - 02 - 12). [2025 - 04 - 18]. https: //ethereum. org/en/whitepaper/.

[2] ZHANG H, WANG C, XU Z, et al. F3B: a low-overhead blockchain architecture with per-transaction front-running protection [J]. arXiv preprint arXiv: 2205. 08529, 2022.

[3] NARAYANAN A, BONNEAU J, FELTEN E. Bitcoin and cryptocurrency technologies [M]. Princeton University Press, 2016.

[4] CHAN T, ZOHAR A. FIFO ordering and its limits in decentralized systems [J]. ACM Transactions on Economics and Computation, 2019, 7 (4): 32 - 50.

[5] ZHU B Z, WAN X, MOALLEMI C C, et al. Quantifying thevalue of revert protection [J]. arXiv preprint arXiv: 2410. 19106, 2024.

[6] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518: 529 - 533.

[7] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of go with deep neural networks and tree search [J]. Nature, 2016, 529: 484 - 489.

[8] ZOHAR A, DAIAN P, JUELS A, et al. Flashbots: a transparent solution for MEV in Ethereum [EB/OL]. (2021 - xx - xx) [2025 - 04 - 18]. https: //flashbots. org.

[9] XU W, ZHANG Y, LIU M, et al. Improving fairness in decentralized exchanges by introducing time-weighted trading priority [J]. Journal of Blockchain Technology, 2021, 14 (2): 85 - 98.

[10] HEIMBACH L, SCHERTENLEIB E, WATTENHOFER R. The potential of selfregulation for frontrunning prevention on DEXes [J]. arXiv preprint arXiv: 2306. 05756, 2023.

[11] SARKAR D. FairFlow protocol: equitable maximal extractable value (MEV) mitigation in Ethereum [J]. arXiv preprint arXiv: 2312. 12654, 2023.

（收稿日期：2025 - 04 - 28）

作者简介：

卢彦胜（1997 - ），男，硕士，主要研究方向：区块链。

李京（1966 - ），男，教授，主要研究方向：组合软件技术、大型网络系统和分布式算法。

（上接第 19 页）

[9] 胡枫杰. 基于 LightGBM 网络入侵检测系统的研究 [D]. 西安：西安电子科技大学，2021.

[10] 孙思佳. 基于模糊 SVD 和极限梯度提升树异常检测方法 [D]. 哈尔滨：哈尔滨工程大学，2021.

[11] KUMAR P, KUSHWAHA C, SETHI D, et al. Investigating the performance of multivariate LSTM models to predict the occurrence of Distributed Denial of Service (DDoS) attack [J]. PLOS ONE, 2025, 20 (1).

[12] 赵玉程，李英建，沈世民，等. 基于网格搜索和投票分类模型的喷油器故障诊断研究 [J]. 机床与液压，2024，52 (5)：213 - 220.

[13] 刘俊泽，汤艳君，薛秋爽. 基于贝叶斯优化 LightGBM 的物联网入侵检测模型 [J]. 警察技术，2022 (5)：73 - 77.

[14] RAJKUMAR K, SHALINIE S M. Semi-supervised deep-ELM for DDoS attack detection and mitigation using the OptimalLink model in IoT networks [J]. Computers & Security, 2025, 152 (C).

[15] SINHA M, BERA P, SATPATHY M, et al. A hybrid lightweight defense system against address spoofing based DDoS attacks in SDN [J]. Security & Privacy, 2025, 8 (2).

（收稿日期：2025 - 05 - 28）

作者简介：

胡宏伟（2000 - ），男，硕士研究生，主要研究方向：计算机网络与信息安全。

孙皓月（1980 - ），男，本科，副教授，主要研究方向：计算机网络与信息安全。

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