

基于样本空间重构的区块链异常检测技术

刘敖迪^{1,2}, 于佳鑫², 杜学绘², 王娜², 曹利峰²

(1. 北方信息技术研究所, 北京 100071; 2. 信息工程大学河南省信息安全重点实验室, 河南 郑州 450000)

摘要: 为了解决区块链网络异常检测难以应对攻击与正常行为动态演化问题, 提出了一种基于样本空间重构的区块链异常检测技术。该方法可识别样本空间的统计分布偏移, 在模型更新中避免灾难性遗忘。该技术的主要思想是首先通过时序自编码器无监督学习正常流量, 构建正常行为的样本空间表示。再通过引入保序回归概率校准异常分数, 使不同时间窗口检测结果具有可比性。采用KL散度、置换检验统计推断流量分布, 识别正常行为导致的样本空间偏移, 区分样本正常变化与异常攻击。最后当检测到偏移后, 自适应更新模型, 在保留正常模式记忆同时, 将新正常样本纳入样本空间。实验结果表明, 区块链异常检测平均AUC值达0.89, 相比于静态模型提升约40%, 误报率降低70%以上。

关键词: 区块链; 流量检测; 样本空间重构; 时序自编码器; 无监督学习

中图分类号: TP309

文献标志码: A

doi: 10.11959/j.issn.1000

Blockchain anomaly detection technology based on sample space reconstruction

Liu Aodi^{1,2}, Yu Jiaxin², Du Xuehui², Wang Na², Cao Lifeng²

1. Northern Institute of Information Technology, Beijing 100071, China

2. He'nan Province Key Laboratory of Information Security, Information Engineering University, Zhengzhou 450000, China

Abstract: To solve the challenge that blockchain network anomaly detection struggles to cope with the dynamic evolution of attacks and normal behaviors, this paper proposes a blockchain anomaly detection technology based on sample space reconstruction. The method can identify the statistical distribution shift in the sample space and avoid catastrophic forgetting during model updates. The core idea of the technique is as follows. First, a temporal autoencoder is used to unsupervised learn normal traffic and construct the sample space representation of normal behaviors. Second, isotonic regression is introduced to probabilistically calibrate anomaly scores, making detection results from different time windows comparable. KL divergence and permutation tests are adopted to statistically infer traffic distributions, identify sample space shifts caused by normal behaviors, and distinguish normal sample variations from anomalous attacks. Finally, once a shift is detected, the model is updated adaptively. While retaining the memory of normal patterns, new normal samples are incorporated into the sample space. Experimental results show that the proposed blockchain anomaly detection method achieves an average AUC of 0.89. It improves performance by approximately 40% compared with static models, and the false positive rate is reduced by more than 70%.

Key words: Blockchain, traffic Detection, sample space reconstruction, temporal autoencoder, unsupervised learning

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 刘敖迪, ladyexue@163.com

基金项目: 国家自然科学基金资助项目(No. 62102449); 河南省重点研发与推广专项(No. 252102211080)

Foundation Items: The National Natural Science Foundation of China (No. 62102449), Key Research and Development and Promotion Program of Henan Province (No. 252102211080)

0 引言

区块链凭借其去中心化、不可篡改等特性,在金融^[1]、医疗^[2]、能源^[3]、物联网^[4]等领域得到广泛应用。作为区块链的通信基础设施,P2P网络承担着交易广播、区块传播、节点发现等核心功能,其安全性和稳定性直接影响整个区块链系统的可靠运行^[5-7]。然而,区块链网络层^[8]正面临着日益严重的攻击威胁,如日食攻击通过控制目标节点的所有入站和出站连接将其与主网隔离^[9],DDoS攻击向目标节点发送大量垃圾数据包耗尽节点资源^[10],交易泛洪攻击通过广播大量无效交易造成网络拥堵^[11]。这些网络攻击的共同特点在于通过操纵、干扰或伪造网络流量来破坏节点间的正常通信过程,能够随着网络环境、协议机制以及防御策略的变化不断调整和升级,这也对区块链异常流量检测的实时性、动态适应性提出了更高要求^[12-14]。

当前针对区块链网络的异常检测方法主要包括基于特征的检测方法、基于时序的检测方法、基于分布的检测方法等。基于特征的检测方法^[15-17]通过提取区块链节点连接状态、交易行为、地址交互行为等多维统计特征,结合机器学习、规则判别等模型构建分类识别机制,以此区分正常节点与恶意攻击行为。该类方法优势在于特征物理意义明确、模型结构简单、训练推理开销小、可解释性强,但高度依赖专家先验知识设计特征,泛化能力较弱,难以识别复杂动态攻击行为。基于时序的检测方法^[18-20]则以区块链流量到达间隔、区块生成时延、节点邻居连接演化、交易请求时序等时序数据为研究对象,捕捉节点流量动态演化规律,实现异常时序行为模式的识别。该类方法的优势是能够更加精准感知、适配网络动态变化场景,但对时序数据的完整性依赖度较高,时序模型参数量大、计算成本较高且可解释性较差。基于分布的检测方法^[21-23]通过拟合区块链正常的网络流量、交易行为及节点交互行为的统计分布规律,构建描述正常行为的基准模型,将偏离基准分布的样本判定为异常攻击行为。该类方法的优势在于无需大量标注样本、具备无监督检测能力、通用性强,但对复杂区块链流量分布拟合难度大,难以精准定位攻击类型与异常成因,检测能力受限。

总的来说,现有面向区块链网络的异常检测时的三大核心挑战。一是攻击手段持续演化,未知隐

蔽威胁识别困难。区块链网络中的攻击模式不断迭代变异,新型攻击手段与攻击变种层出不穷。现有检测方法多依赖已知攻击特征与监督训练范式,只能拟合已标注攻击样本模式,缺乏对攻击的泛化识别能力,难以满足攻击手段动态演化带来的异常检测需求。二是网络行为动态偏移,导致模型性能衰退。区块链网络的节点规模、交易频次、网络拓扑随时间持续动态变化,正常业务行为存在时序漂移与分布偏移,难以用固定不变的行为基准表达。传统检测多采用静态建模方式,模型一旦训练完成便难以跟随正常行为实时迭代更新,无法动态适配样本分布的持续偏移,极易引发漏报、误报问题。三是标注样本稀缺,监督检测模型落地受限。区块链恶意流量精准标注需要专业领域知识与大量人工核验,标注成本高、难度大,难以获取足量高质量标注数据集。而现有检测高度依赖带标签样本支撑模型训练,受限于标注数据稀缺瓶颈,模型训练不充分、泛化性差,实际应用的适用性严重受限。

针对以上挑战,本文提出一种基于样本空间重构的区块链异常检测框架BADT-SSR,将网络行为的偏移感知机制引入区块链异常检测,通过无监督的时序自编码器构建样本空间、保序回归实现行为概率校准、KL散度与置换检验检测分布偏移、弹性权重固化完成模型更新,从而实现动态、自适应的区块链异常检测。主要贡献包括5个方面:

(1) 提出面向区块链异常检测的样本空间重构机制。构建涵盖“样本空间构建、概率校准、偏移检测、自适应更新”的样本空间重构闭环,实现漂移检出与持续学习更新的受控联动。

(2) 提出基于时序自编码器的样本空间构建方法。通过时序自编码器无监督学习历史正常流量,将区块链流量时序特征编码为可度量的重建误差,构建正常行为的样本空间表示,为后续偏移检测奠定特征基础。

(3) 提出基于保序回归的样本空间概率校准方法。引入保序回归概率校准异常分数,将原始分数映射为误报率FPR,使不同时间窗口的检测结果具有可比性,为正常与异常流量分布偏移的检测提供可靠输入。

(4) 提出基于KL散度与置换检验的样本空间偏移检测方法。采用KL散度量化分布差异,结合置换检验进行显著性判断,识别出区块链网络产生

的样本空间统计偏移。

(5) 提出基于弹性权重的样本空间自适应更新方法。当检测到偏移后,通过弹性权重机制自适应更新异常流量检测模型,在保留原有正常模式记忆同时,将新样本纳入样本空间,解决灾难性遗忘问题,确保异常流量检测标准与动态环境同步。

1 相关工作

在区块链网络环境下,异常检测已成为得到广泛关注的研究重点。按照行为建模视角与异常判定机理不同,现有主流检测方法可归纳为基于特征的检测方法、基于时序的检测方法、基于分布的检测方法三类范式。三类方法分别在特征表征、时序依赖、分布拟合等不同层面具有优势,但在泛化性、实时性与灵活性也存在各自局限。

在基于特征的检测方面,戴千一等^[15]提出一种基于多分类器集成的区块链网络层异常流量检测方法,通过区分度和冗余信息量实现流量特征子集选择,确保特征提取质量。Su等^[16]提出一种基于eBPF和大模型的区块链异常检测方案,将用于内核级安全监控的扩展伯克利包过滤技术与大模型结合,集成用于上下文区块链威胁情报的检索增强生成和多步攻击分析思维链提示,识别潜在异常行为。针对模型构建中存在的隐私信息泄露风险,Li等^[17]提出一种分布式入侵检测系统,采用异步联邦学习算法,通过有向无环图实现客户端节点模型参数可信共享,利用监控链检测区块链网络层中的DDoS攻击。针对现有方法提取特征时面临内存消耗高、检测精度低的挑战,朱龙隆等^[24]提出一种基于近似测量算法Sketch的区块链异常检测机制,将区块链异常检测转化为异常交易账户检测,使用Sketch算法识别账户,通过聚合和分析多区块信息实现账户准确检测。Voronov等^[25]也提出一种基于Sketch的区块链网络异常检测框架,利用Bloom Filter和HyperLogLog,在不需要检查整个区块链数据时高效识别异常账户。Shen等^[26]提出基于图的区块链异常检测持续学习方法,通过权重保持模块捕获局部结构依赖,在训练过程中稳定关键参数,通过识别建模拓扑重要性,减轻灾难性遗忘。Wang等^[27]提出一种最小的替换标记传播模型MSLP,提供更多标记数据平衡图形数据,补充区块链网络中异常事务检测服务样本,利用社会计算

领域最小替代理论,实现少量计算成本下发现异常交易。总的来说,基于特征的检测依赖专家先验知识,仅能识别浅层、规则化攻击行为,难以刻画动态隐蔽攻击特征,泛化能力有限,易产生漏检与误检。

在基于时序的检测方面,Liu等^[18]提出一种基于图注意力网络的区块链异常检测方法,采用动态属性图构建策略,以边关系对每笔交易进行结构化建模,通过学习区块链网络的时序演化规律,动态更新不同时间切片下子图的节点学习权重,充分挖掘区块链节点的深层隐含特征,并采用二分类模型完成区块链网络节点的异常识别。Chen等^[19]提出一种用于区块链异常检测的多距离时空图神经网络模型MDST-GNN,将多距离图卷积架构与自适应时间建模集成,结合自监督学习,捕获本地和全局的时空依赖关系,支持从匿名时间数据中推断异常模式,增强了模型泛化能力。Du等^[20]提出一种基于动态图学习的细粒度类增量恶意账户检测框架FiMAD,通过动态账户关系图从连续时间角度对以太坊账户动态建模,设计了一种级联图特征提取方法捕获交互特征,通过预训练通用编码器将账户特征转换为高维嵌入,使用少量标记样本对模型分类器微调,实现精确的细粒度检测和增量类快速更新。Liu等^[28]从智能合约中引入不同交易类型挖掘账户本身固有的时间交易模式入手,将每个以太坊账户的多个异构交易间的隐式转换关系建模为异构的、时间的、有向的交易状态转换图,利用基于注意力的图神经网络实现对状态转移图的学习和以太坊网络的钓鱼检测。总的来说,基于时序的检测对完整、高质量时序数据依赖性强,且抗噪声能力较弱,难以识别低强度、慢渗透的隐蔽攻击。同时,模型计算开销大、可解释性差,对区块链渐进式、复杂化的异常行为识别存在性能瓶颈。

在基于分布的检测方面,Li等^[21]提出一种基于对比学习和关系机制的异常行为检测方法,利用区块链中大量未标记的数据及各种事务节点特征间关系,探索交易节点的特征分布,通过对比学习推断节点簇的标签属性,利用关系机制构建节点在事务过程的行为和结构特征信息,为节点异常检测提供全面的理解和支持。针对不同的以太坊行为在加密和持久连接传输难以区分的问题,Hu等^[22]提出一种细粒度的以太坊行为识别系统,通过序列化向

后推理的加密流量分析,有效地识别以太坊网络上的异常行为,实现对以太坊用户的有效监管。Kim等^[23]提出一种基于半监督学习的区块链网络流量分析机制,该机制由数据收集和异常检测两个模块组成,数据采集模块引擎通过感知底层区块链流量,周期性、实时地生成多维数据流,而异常检测模块从创建的正常数据实例构建正常行为为基线以检测异常攻击流量,具备一定检测未知攻击模式的能力。针对 Execute-Order-Validate 区块链通过允许并行事务执行来增强性能,但是它也引入了可能导致分类帐中状态不一致的事务冲突,Zhang等^[29]分别研究了在泊松分布、高斯分布、未知分布等不同区块链分组间隔分布下,操纵事务冲突攻击的攻击效率情况,并采用随机森林算法实现对该类攻击的有效检测。但总的来说,基于分布的检测无需标注数据、通用性较强,但仅能判断行为整体异常偏离程度,难以精准区分攻击类型、定位异常成因,且难以拟合区块链复杂的行为分布,细粒度检测能力存在不足。

上述研究主要面向区块链领域的异常识别任务展开,较少结合流式数据分析领域的通用概念漂移检测、持续学习相关成果开展交叉研究。为进一步厘清本方法的研究定位,下面对通用概念漂移检测与通用持续学习领域的代表性工作进行梳理。概念漂移^[30-31]指数据流中数据的联合分布随时间发生变化,是流式数据分析领域广泛研究的问题。现有概念漂移检测算法可分为基于统计检验的漂移检测、基于窗口的漂移检测、基于模型预测误差监控的漂移检测^[32]。但现有通用漂移检测框架存在如下不足:一是多数检测方法直接输出漂移告警,缺少告警甄别环节,流量随机噪声极易产生虚假漂移告警。二是大部分漂移检测工作聚焦“如何检测漂移”,不考虑漂移之后模型如何受控更新,漂移检测模块与持续学习模块割裂。一旦检测漂移就执行模型更新,虚假漂移会污染历史模型记忆。此外,很多漂移检测直接使用原始特征或者原始模型输出分数,没有前置分数标尺对齐步骤,原始输出随时间发生量纲漂移,会直接降低漂移检测可靠性。而持续学习^[33-34]致力于解决模型在顺序任务学习过程中的灾难性遗忘问题,主流方案包含正则化类、知识蒸馏类、动态架构类、回放类方法。但现有持续学习算法大多聚焦优化更新过程本身,常常采用固

定周期更新策略,存在把异常样本、噪声吸收进历史正常模式,造成模型退化的问题。

2 区块链异常检测框架

2.1 框架概述

基于样本空间重构的区块链异常检测框架 BADT-SSR 如图 1 所示,由样本空间构建、概率校准、偏移检测、自适应更新四个核心模块组成,形成从正常模式学习到动态自适应模式更新的闭环。

(1) 样本空间构建模块:通过时序自编码器对历史正常流量进行无监督学习,将交易广播、区块同步等区块链典型行为的时序特征编码为可度量的重建误差,由此构建正常行为的样本空间表示。这一阶段输出初始模型和控制集,后者记录了历史正常样本的特征分布,作为后续比较的基准。

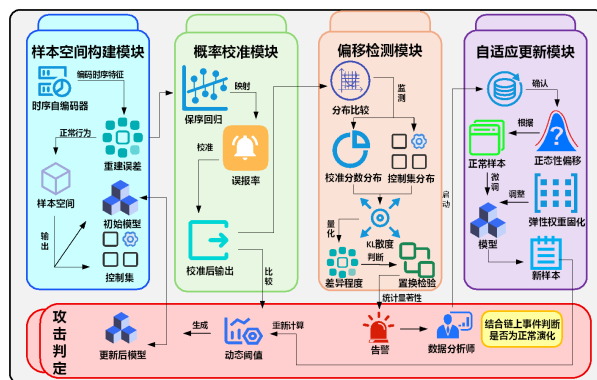


图1 BADT-SSR 异常检测框架

(2) 概率校准模块:接收样本空间构建输出的原始重建误差,通过保序回归将这些分数映射为误报率,使校准后输出落在0到1之间且具有统一的量纲。校准后的分数同时流向两个方向,一路进入攻击判定环节,与动态阈值比较以输出检测结果。另一路进入偏移检测模块,作为分布比较依据。

(3) 偏移检测模块:持续监测当前窗口的校准分数分布与控制集分布间差异,采用KL散度量差异程度,借助置换检验判断差异是否具有统计显著性。若p值小于0.05,系统触发告警,结合链上事件确认是否为正常演化。确认为正常变化时启动自适应更新模块,否则维持现有模型参数。

(4) 自适应更新模块:在确认当前偏移为正常行为发生变化导致的之后,利用当前窗口的正常样本依托弹性权重固化机制对模型进行微调。弹性权重固化机制通过Fisher信息矩阵度量参数重要性,

在更新时对重要参数施加惩罚,从而在适应新分布的同时避免遗忘原有模式的正常模式。微调完成后,新样本加入控制集,动态阈值重新计算,更新后的模型将被反馈回样本空间构建模块用于下一轮的异常检测。图 1 自适应更新模块标注的“正态性偏移”指代上游样本空间偏移检测模块输出的、经过多级确认流程判定为正常业务演化引发的样本空间统计偏移事件信号。该信号是触发 EWC 自适应微调的前置输入条件,本身不执行任何模型运算。只有接收到该确认后的正常偏移信号,自适应更新模块才会启动后续的 EWC 微调流程。若无该偏移触发信号,自适应更新模块保持休眠,不会执行模型微调,以此避免无意义的频繁更新。

攻击判定环节贯穿整个检测流程,但不作为独立模块存在。动态阈值取当前控制集校准分数的 95% 分位数,新样本的校准分数高于阈值则判为攻击,否则判为正常行为。模型更新后阈值随之重新计算,确保检测标准与环境同步。总的来说,样本空间构建提供基础表示,概率校准使分数具备可比性,偏移检测识别分布变化并决定是否更新,自适应更新将新知识融入模型并反馈至起点,攻击判定则始终以动态阈值完成检测。

2.2 样本空间构建

(1) 面向区块链网络的特征工程

从原始流量中提取有效的特征是区块链异常检测的基础。针对区块链网络的特殊性,从以下四个维度,全面刻画区块链网络会话行为模式。

(1) 流量特征描述网络会话基本数量属性和动态变化,包括 8 个特征,如表 1 所示。

表 1 流量特征集	
特征	含义
Total_bytes	会话传输的总字节数
Total_packets	会话包含数据包总数
Duration	会话的持续时间(秒)
Packets_per_second	每秒传输的数据包数
Bytes_per_packet	平均数据包大小
Byte_rate_variance	字节速率的方差
Packet_size_entropy	数据包大小分布的熵
Arrival_time_mean	数据包到达间隔均值

(2) 连接特征刻画区块链节点之间的交互模式和 IP 多样性,对日食攻击等连接级攻击敏感,包括 6 个特征,如表 2 所示。

表 2 连接特征集	
特征	含义
In_connections	入站连接数
Out_connections	出站连接数
Con_success_rate	连接成功率
Src_ip_diversity	源 IP 多样性
Dst_ip_diversity	目标 IP 多样性
Src_ip_entropy	源 IP 分布熵

(3) 协议特征描述会话使用的网络协议分布,反映攻击者可能利用的协议异常,包括 5 个特征,如表 3 所示。

表 3 协议特征集	
特征	含义
Is_tcp	是否使用该协议
Is_udp	是否使用该协议
Is_icmp	是否使用该协议
Is_dns	是否使用该协议
Protocol_entropy	协议分布的熵

(4) 节点状态特征描述节点的运行健康状态,反映节点是否被攻陷或孤立,包括 5 个特征,如表 4 所示。

表 4 节点状态特征集	
特征	含义
Cpu_usage	CPU 使用率(%)
Memory_usage	内存使用率(%)
Node_uptime	节点连续运行时间
Peer_count	对等节点数量
Peer_churn_rate	对等节点变更率

对于每个区块链网络会话,将上述 24 个特征组合成特征向量 $\mathbf{x} \in \mathbf{R}^{24}$ 。对于时间窗口内的多个区块链网络会话,按时间顺序排列不同会话特征,构成时序特征序列 $\mathbf{X}=[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]$,作为后续模型的输入,其中 T 为序列长度。在真实区块链场景中,考虑到网络会话的持续时间变化,我们按照常用设

置, 采用固定长度 $T=10$, 长度不足的序列进行特征填充, 长度超过的则截断。

(2) 时序自编码器与样本空间表示

时序自编码器由编码器和解码器两部分组成。

(1) 编码器

输入序列 $X=[x_1, x_2, \dots, x_T]$ 首先通过第一层 LSTM 网络, 得到每个时间步的隐藏状态 $h_t^{(1)} \in R^H$:

$$h_t^{(1)} = LSTM_1(x_t, h_{(t-1)}^{(1)})$$

其中, H 为隐藏层维度, $h_t^{(1)}$ 为第一层返回每个时间步的隐藏状态。再将 $h_t^{(1)}$ 输入第二层 LSTM 网络, 得到第二层隐藏状态:

$$h_t^{(2)} = LSTM_2(h_t^{(1)}, h_{(t-1)}^{(2)})$$

第二层 LSTM 网络仅返回最后一步的隐藏状态 $h_t^{(2)}$ 作为整个序列的压缩表示, 即上下文向量 $c = h_T^{(2)} \in R^H$ 。

(2) 解码器

解码器以上下文向量 c 作为初始状态, 逐步重建输入序列。将 c 复制 T 次得到 $[c, c, \dots, c]$, 作为解码器的输入序列:

$$\tilde{h}_t^{(1)} = LSTM_1^{dec}(c, \tilde{h}_{(t-1)}^{(1)})$$

然后, 输入第二层 LSTM 网络解码:

$$\tilde{h}_t^{(2)} = LSTM_2^{dec}(\tilde{h}_t^{(1)}, \tilde{h}_{(t-1)}^{(2)})$$

最后通过全连接层输出重建序列:

$$\hat{x}_t = W_o \tilde{h}_t^{(2)} + b_o$$

其中, $W_o \in R^{(24 \times H)}$, $b_o \in R^{24}$ 为可训练参数。解码器输出为重建序列 $\hat{X} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_T]$ 。

(3) 异常分数

对于输入序列 X , 模型的重建误差定义为每个时间步特征向量的均方误差的平均值:

$$s(X) = \frac{1}{T} \sum_{t=1}^T \frac{1}{24} \sum_{j=1}^{24} (x_{tj} - \hat{x}_{tj})^2$$

其中, x_{tj} 为第 t 个时间步的第 j 维特征, $\hat{x}_{tj}$ 为对应的重建值。 $s(X)$ 即为样本 X 的异常分数。

模型训练时, 仅使用正常样本, 优化目标是最小化重建误差:

$$L_{AE} = \frac{1}{N} \sum_{i=1}^N s(X_i)$$

其中, N 为训练样本数。在测试时, 异常样本由于模式未被模型学习, 重建误差通常较大。

样本空间构建模块负责将区块链网络流量的时序特征编码为可度量的样本空间表示。该模块采用

Algorithm 1: Training of Temporal Autoencoder

INPUT: Training set $D_{\text{train}} = \{X_1, X_2, \dots, X_N\}$, Hidden dimension H , Sequence length T , Number of epochs E , Batch size B

OUTPUT: Trained model parameters θ

```

1: Initialize encoder and decoder parameters  $\theta_{\text{enc}}, \theta_{\text{dec}}$ 
2: FOR epoch = 1 to E DO
3:   Shuffle  $D_{\text{train}}$ 
4:   FOR each mini-batch  $\{X_1, \dots, X_B\}$  in  $D_{\text{train}}$  DO
5:     FOR each sequence  $X$  in the batch DO
6:       Compute hidden states  $h_t^{(1)}$ 
7:       Compute hidden states  $h_t^{(2)}$ 
8:       Context vector  $c = h_T^{(2)}$ 
9:       Reconstruct  $\tilde{h}_t^{(1)}, \tilde{h}_t^{(2)}, \hat{x}_t$ 
10:    END FOR
11:   Compute reconstruction loss  $L$ 
12:   Backpropagate and update  $\theta_{\text{enc}}, \theta_{\text{d}}$ 
13: END FOR
14: END FOR
15: RETURN  $\theta$ 

```

时序自编码器对历史的正常流量进行无监督学习, 通过重建误差来量化当前样本偏离正常模式的程度。在训练过程中, 其仅使用正常样本, 使模型能够准确编码交易广播的突发性、区块传播的周期性等区块链典型行为的时序特征。训练完成后, 模型参数构成了正常行为的样本空间表示, 输出初始控制集 (历史正常样本的异常分数分布) 作为后续概率校准和偏移检测的基准。时序自编码器训练具体流程如算法 1 所示。

2.3 样本空间概率校准

(1) 目标校准

直接使用时序自编码器输出的原始重建误差 (原始异常分数) 开展跨时间窗口分布对比存在两处关键缺陷: 一是原始重建误差取值范围随流量分布动态变化, 不同时间窗口输出的原始分数量纲不一致, 不具备横向可比性; 二是原始重建误差仅表征样本与历史正常模式的距离, 不存在明确的概率语义, 难以用于统计显著性判定与动态阈值设定。为此, BADT-SSR 框架引入保序回归概率校准, 对无监督模型输出的原始异常分数做后处理变换, 将无概率语义的重建误差映射至 $[0, 1]$ 区间, 输出具备经验误报率 FPR 物理含义的校准异常分数。

保序回归概率校准: 给定仅包含正常样本的训练集及其对应的原始异常分数集合, 保序回归概率校准是一类非参数单调拟合过程。首先依据原始异常分数大小对正常样本升序排序, 计算每个样本对应的经验误报率 FPR ; 再利用保序回归拟合得到单

调非递增映射函数 $C(\cdot)$, 该函数将任意输入原始异常分数 s 映射为校准分数 $C(s) \in [0, 1]$ 。该映射严格保持原始异常分数的相对大小顺序, 即若 $s_1 < s_2$, 则 $C(s_1) \geq C(s_2)$ 。

对于训练集内全部正常样本集合 $X_N = \{X_1, X_2, \dots, X_n\}$, 共 n 个正常样本, 计算每个样本的原始异常重建误差 $s_i = s(X_i)$ 。将全部正常样本按照原始异常分数从小到大升序排序, 得到有序序列:

$$s_1 \leq s_2 \leq \dots \leq s_i \leq \dots \leq s_n$$

其中, 下标 $i=1, 2, \dots, n$ 为排序之后样本序号。

对排序后第 i 个正常样本, 经验误报率 FPR 表示: 若选取该样本原始异常分数 s 作为检测阈值时, 训练集中会被误判为异常的正常样本占比。计算公式为:

$$FPR_{T=s}(x_N) = \frac{|\{x \in X_N : s(x) > s\}|}{x_N}$$

当原始异常分数 s 作为判定阈值时, 所有原始分数大于 s 的正常样本会被误报。排序后序号越大 (原始异常分数越高), 对应的经验 FPR 越小。

保序回归拟合得到单调非递增校准函数 $C(s)$ 。对于任意输入原始异常分数 s , 输出校准分数 $C(s)$, 其物理含义为当以原始异常分数 s 作为检测阈值时, 历史正常样本集合中会被误判为异常的样本比例。

$C(s)$ 数值越接近 1, 代表该原始分数下会产生大量误报, 对应样本偏离正常模式程度越低, 越倾向正常; 而 $C(s)$ 数值越接近 0, 代表该原始分数下几乎不会产生误报, 对应样本显著偏离历史正常模式, 越倾向攻击异常。

假设正常样本训练集样本总数 $n=10$, 10 个正常样本经时序自编码器得到原始重建误差 (原始异常分数): $[0.12, 0.15, 0.18, 0.21, 0.26, 0.33, 0.41, 0.47, 0.50, 0.62]$, 原始重建误差转换校准分数过程示例如下:

1) 升序排序: 样本已经从小到大排列, 排序后索引 i 依次取 $1, 2, \dots, 10$;

2) 计算每个样本对应的经验 FPR :

$$i=1, s_1=0.12, FPR_1=0.9;$$

$$i=2, s_2=0.15, FPR_2=0.8;$$

$$i=3, s_3=0.18, FPR_3=0.7;$$

$$i=4, s_4=0.21, FPR_4=0.6;$$

$$i=5, s_5=0.26, FPR_5=0.5;$$

$$i=6, s_6=0.33, FPR_6=0.4;$$

$$i=7, s_7=0.41, FPR_7=0.3;$$

$$i=8, s_8=0.47, FPR_8=0.2;$$

$$i=9, s_9=0.50, FPR_9=0.1;$$

$$i=10, s_{10}=0.62, FPR_{10}=0.0.$$

3) 将 $\{s_i, FPR_i\}$ 成对输入保序回归, 拟合单调非递增函数 $C(s)$ 。

4) 对于新样本, 原始重建误差 $s=0.50$, 输入校准函数得到校准输出 $C(0.50)=0.1$ 。该结果含义为当原始异常分数阈值取 0.50 时, 训练集内仅 10% 正常样本会被误判异常; 该样本相比 90% 历史正常样本偏离程度更大, 具备较高异常嫌疑。

(2) 校准器拟合

对训练集正常样本的异常分数进行升序排序, 得到序列 $a_1 \leq a_2 \leq \dots \leq a_n$, 其中 $n=|x_N|$ 。

对每个 a_i , 计算对应的经验 FPR :

$$a_i = \frac{(n-i)}{n}$$

使用保序回归拟合单调非递增函数 $C: a_i \mapsto b_i$, 约束条件为 $C(a_i) \geq C(a_{i+1})$ 。

保序回归是一种非参数回归方法, 能够在不预设函数形式的前提下, 找到满足单调性的最优拟合函数。它通过求解以下二次规划问题实现:

$$\min_C \sum_{i=1}^n (C(a_i) - b_i)^2, s.t. C(a_i) \geq C(a_{i+1})$$

该问题保证校准函数单调非递增且自由形式, 能够拟合复杂的分布。

对于新的异常分数 s , 校准输出为 $\hat{s}$, 确保落在 $[0, 1]$ 区间:

$$\hat{s} = \min(1, \max(0, C(s)))$$

样本空间概率校准具体流程如算法 2 所示。本文保序回归校准仅利用初始训练期的历史正常样本一次性拟合得到, 在后续全部时间窗口内固定不再重新训练、不再重拟合。若每个时间窗口利用当前窗口样本重新拟合校准, 校准会吸收当前窗口的分布漂移信息, 将真实业务演化带来的分布差异抹除, 导致后续 KL-置换偏移检测模块无法识别样本空间偏移。因此本框架采用固定校准器策略, 既完成跨窗口分数标尺对齐, 又保留原始数据中真实存在的分布结构差异, 保障后续偏移检测可以正常工作。

Algorithm 2: Probability Calibration via Isotonic Regression

INPUT: Raw anomaly scores $S=s_1, s_2, \dots, s_n$, from normal training set, Corresponding empirical FPR values $b_i = (n-i)/n$
OUTPUT: Calibration function $C: [0, \infty) \rightarrow [0, 1]$
 1: Sort S in ascending order to obtain $a_1 \leq a_2 \leq \dots \leq a_n$
 2: Compute $b_i = (n-i)/n$ for $i = 1..n$
 3: Use isotonic regression to fit monotonic non-increasing function C
 4: FOR new score s , define calibrated score as $\hat{s} = \min(1, \max(0, C(s)))$
 5: RETURN calibration function C

2.4 样本空间偏移检测

为了区分区块链网络的正常行为演化（如协议升级、新应用上线）与真正的攻击，BADT-SSR 框架实时监测流量分布是否发生显著变化。通过比较当前窗口与控制集的校准分数分布，采用 KL 散度结合置换检验进行统计推断。值得注意，本模块使用 KL 散度结合置换检验与概率校准评估阶段所用 KS 检验定位不同。KS 检验用于验证校准后正常样本分数的跨窗口可比性，本处检验是在统一可比的分数标尺之上识别样本集合真实分布结构差异。

假设控制集 D_c 为历史正常样本的校准分数集合，处理集 D_t 为当前时间窗口所有样本的校准分数集合（可能包含攻击和正常变化）。分布偏移检测的目标是判断 D_c 和 D_t 是否来自同一分布。将 $[0, 1]$ 区间划分为 K 个等宽区间，即校准分箱数，统计控制集和处理集样本落在各区间的频率，得到离散概率分布 P 和 Q ：

$$P(i) = \frac{|\{s \in D_c : s \in \text{bin}_i\}|}{|D_c|}$$

$$Q(i) = \frac{|\{s \in D_t : s \in \text{bin}_i\}|}{|D_t|}$$

为避免零概率问题，对每个区间计数加 1 后再归一化（Laplace 平滑）。

计算离散概率分布 P 和 Q 的 KL 散度：

$$D_{KL}(P||Q) = \sum_{i=1}^K P(i) \log \frac{P(i)}{Q(i)}$$

KL 散度越大，表示两个分布差异越大。KL 散度具有非负性，且当且仅当 $P=Q$ 时为零。

为了评估观测到的 KL 散度是否具有统计显著性，进行置换检验，将控制集和处理集的分数进行合并，得到合并后的集合 $S = D_c \cup D_t$ ，大小为 $|D_c| + |D_t|$ 。随机打乱 S ，将其前 $|D_c|$ 个作为置换控制集 D'_c ，剩余为置换处理集 D'_t 。计算置换 KL 散度：

$$D_{KL}^{(i)} = D_{KL}(P' || Q')$$

其中， P' 和 Q' 分别是基于 D'_c 和 D'_t 的频率分布。计算 p 值：

$$p = \frac{1 + \#\{i : D_{KL}^{(i)} \geq D_{KL}^{(obs)}\}}{N_{perm} + 1}$$

其中， $D_{KL}^{(obs)}$ 为原始观测到的 KL 散度。若 $p < 0.05$ ，则拒绝原假设即两组来自同一分布，认为发生了显著偏移。

Algorithm 3: Distribution Shift Detection

INPUT: Control set calibrated scores D_c , Treatment set calibrated scores D_t , Number of bins K , Number of permutations N_{perm} , Significance level $\alpha = 0.05$
OUTPUT: $\text{shift}_{\text{flag}}$ (True if shift detected, False otherwise), p_{value}
 1: Partition $[0, 1]$ into K equal-width bins
 2: Compute histograms P from D_c and Q from D_t
 3: Compute observed KL divergence $D_{KL}^{(obs)}$
 4: Concatenate $D = D_c \cup D_t$
 5: $n_c = |D_c|$, $n_t = |D_t|$, count = 0
 6: FOR $i = 1$ to N_{perm} DO
 7: Randomly shuffle D
 8: Take first n_c elements as permuted control D'_c , remaining as D'_t
 9: Compute histograms P' from D'_c and Q' from D'_t
 10: Compute permuted KL divergence D_{perm}
 11: IF $D_{perm} \geq D_{obs}$ then
 12: count = count + 1
 13: END IF
 14: END FOR
 15: $p_{\text{value}} = (\text{count} + 1) / (N_{perm} + 1)$
 16: $\text{shift}_{\text{flag}} = (p_{\text{value}} < \alpha)$
 17: RETURN $\text{shift}_{\text{flag}}$, p_{value}

当检测到显著偏移时，系统触发告警，提示系统关注当前窗口的数据分布变化。该告警仅作为变化提示信号，不能直接驱动模型更新，告警需要经过自动初筛、链上事件匹配、人工复核多级确认流程。只有确认偏移来源于协议升级、新业务上线等正常业务演化，才触发自适应更新模块。若判定为随机流量波动、网络攻击事件，或者无法确认偏移来源，则维持现有模型并加强监控，避免因流量偶然波动导致的误更新。偏移告警完整确认流程见 3.3 节的部署应用讨论内容。样本空间偏移检测具体流程如算法 3 所示。

为更加准确的描述样本空间自适应更新触发条件，偏移告警到自适应更新触发的形式化定义如下：

设 $\text{Flag}_{\text{shift}}$ 表示 KL-置换偏移检测输出布尔标志， $\text{Flag}_{\text{shift}} = \text{True}$ 代表检测到统计分布偏移告警，False 无偏移告警。 S_{alert} 表示偏移告警窗口内全部

会话样本集合。 $Flag_{filter}$ 表示自动初筛模块输出的布尔标志。 $Flag_{filter}=True$ 代表经过阈值、样本量、偏移幅度过滤之后,告警通过自动初筛, $False$ 代表告警被自动初筛判定为流量随机波动,直接丢弃告警。 E_{chain} 表示链上事件匹配输出事件集合,记录该时间窗口发生的协议升级、新应用上线等真实业务演化事件, $E_{chain}=\emptyset$ 表示无匹配业务事件。 $Flag_{human}$ 表示运维人员人工复核输出布尔标志, $Flag_{human}=True$ 代表人工确认该偏移来源于正常业务演化, $False$ 代表判定为攻击/噪声,拒绝触发更新。

自适应更新触发的充分必要条件为:

$$Flag_{update} = Flag_{shift} \wedge Flag_{filter} \wedge (E_{chain} \neq \emptyset) \wedge Flag_{human}$$

仅当 $Flag_{update}=True$ 时,才允许将当前窗口经过筛选的正常样本送入执行模型自适应更新,只要任意一项条件不满足,则 $Flag_{update}=False$,禁止执行模型更新,维持原有模型、控制集、动态阈值不变。从该形式化定义可以看出,仅仅偏移检测输出告警 $Flag_{shift}=True$ 不足以触发更新,必须同时满足自动初筛通过、存在对应的链上业务事件佐证、人工复核确认为正常业务演化三个附加约束。该多重约束也是抵御攻击者刻意诱导模型更新的重要屏障。

2.5 样本空间自适应更新

当检测到样本空间偏移且经确认属于正常变化后,需要对模型进行更新,以适应当前的正常模式。同时,我们希望模型能够保留之前学习到的正常知识,避免因适应新的样本数据而遗忘旧的知识。BADT-SSR 框架采用弹性权重固化 (Elastic Weight Consolidation, EWC) 解决这一问题。

(1) 弹性权重固化

弹性权重固化 EWC 是对于 Fisher 信息 F_i 较大的参数 (即对旧知识贡献大的参数),任何与旧值 $\theta_{old,i}$ 的偏差都会受到较大惩罚,因此这些参数在更新时变化较小,保留了旧任务的记忆。而对于 F_i 较小的参数 (即对旧知识贡献小的参数),惩罚较小,可以灵活调整以适应新任务。这种机制在适应新分布的同时有效缓解了灾难性遗忘问题。

EWC 的核心思想是在贝叶斯框架下,参数 θ 对于旧任务 (历史数据) 的重要性可以通过 Fisher 信息矩阵来度量。在优化新任务时,对重要参数的更

新施加惩罚,从而在适应新分布的同时保留旧知识。设旧任务 (历史正常数据) 的参数后验分布为 $p(\theta|D_{old})$,新任务 (当前正常数据) 的对数似然为 $\log p(D_{new}|\theta)$ 。根据拉普拉斯近似原理, $p(\theta|D_{old})$ 可以近似为以 θ_{old}^* 为中心的高斯分布,协方差矩阵的逆为 Fisher 信息矩阵 F 。因此,新任务的优化目标为:

$$L(\theta) = L_{new}(\theta) + \lambda \sum_i \frac{F_i}{2} (\theta_i - \theta_{old,i})^2$$

其中, L_{new} 是新任务的损失函数 (重建误差), λ 是正则化强度, F_i 是 Fisher 信息矩阵的对角元素,度量参数 θ_i 对旧任务的重要性。

(2) 重要性计算

在 BADT-SSR 框架中,旧任务对应于训练集和控制集的正常样本,新任务对应于当前窗口中经确认的正常样本。参数重要性 F_i 的计算方式为对旧任务中的每个样本 x ,计算损失函数 (即重建误差):

$$l(x) = \frac{1}{24} \sum_{j=1}^{24} (x_j - x^j)^2$$

计算损失对参数 θ_i 的梯度:

$$g_i(x) = \frac{\partial l(x)}{\partial \theta_i}$$

Fisher 信息近似为梯度的平方期望:

$$F_i = \frac{1}{N_{old}} \sum_{x \in D_{old}} g_i(x)^2$$

其中, N_{old} 为旧任务样本数。计算 Fisher 信息时,冻结模型参数是指关闭优化器的参数更新步骤 (不执行梯度下降权重更新),模型参数保持原有数值不变,但允许正常执行前向传播与反向传播运算求解梯度,并非禁止梯度计算。

该近似存在明确概率假设:一是样本之间满足独立同分布。二是自编码器重建损失采用均方误差,等价于对输出特征施加高斯噪声假设,重建损失对应负对数高斯似然。在上述假设下,损失对参数梯度平方的样本期望,可以作为 Fisher 的无偏近似估计。计算 Fisher 信息时遍历旧任务样本,累积梯度平方后取样本平均,只求解梯度,不执行参数权重更新。

(3) 自适应训练流程

需要更新模型时,首先准备数据,从当前窗口中提取正常样本,构成新任务数据集 D_{new} 。然后计算 Fisher 信息,冻结模型参数,使用旧任务数据集

D_{old} 计算 Fisher 信息矩阵 F 。之后解冻模型，恢复参数可训练状态。最后，使用新任务数据集进行微调，优化目标为：

$$L(\theta) = L_{AE}(\theta; D_{new}) + \lambda \sum_i \frac{F_i}{2} (\theta_i - \theta_{old,i})^2$$

其中， L_{AE} 是重建误差损失， θ_{old} 是旧参数。将当前窗口的正常样本加入控制集（可设置最大容量，超出时丢弃最旧样本）。基于更新后的控制集校准分数，重新计算动态阈值。

样本空间自适应更新具体流程如算法 4 所示。

Algorithm 4: Adaptive Update

INPUT: Old model parameters θ_{old} , Old task dataset D_{old} , New task dataset D_{new} , Regularization strength λ , Number of fine-tuning epochs E , Batch size B

OUTPUT: Updated model parameters θ

// Step 1: Compute Fisher Information Matrix

1: Freeze model parameters at θ_{old}

2: Initialize Fisher diagonals $F_i = 0$ for each parameter i

3: FOR each sample x in D_{old} DO

4: Compute reconstruction loss $l(x)$

5: Compute gradient $g_i(x)$ for each parameter i

6: Accumulate squared gradient: F_i

7: END FOR

8: $F_i = F_i / |D_{old}|$ // average over all samples

// Step 2: Fine-tune on new data with EWC penalty

9: Unfreeze model parameters, initialize $\theta = \theta_{old}$

10: FOR epoch = 1 to E DO

11: FOR each mini-batch X_{batch} in D_{new} DO

12: Compute reconstruction loss L_{new}

13: Compute EWC regularization term L_{ewc}

14: Total loss $L = L_{new} + L_{ewc}$

15: Backpropagate and update θ

16: END FOR

17: END FOR

18: RETURN θ

2.6 攻击判定

BADT-SSR 框架采用动态阈值进行攻击判定，阈值随当前正常分布自动调整，避免静态阈值导致的误报。设当前控制集 D_c 的校准分数集合为 $\{s^1, s^2, \dots, s^m\}$ 。选择 q 分位数作为阈值：

$$\tau = \text{Percentile}_q(\{ \hat{s} : s \in D_c \})$$

即阈值设定为使得 95% 的正常样本分数低于该值。这意味着正常情况下，最多有 5% 的正常样本会被误判为异常，误报率理论上能够控制在 5% 以内。对于新样本 X ，其异常分数为 $s(X)$ ，校准后得到 $\hat{s} = C(s(X))$ 判定规则为：

$$\text{label} = \begin{cases} \text{攻击}, & \text{if } \hat{s} < \tau \\ \text{正常}, & \text{otherwise} \end{cases}$$

当模型自适应更新后，控制集分布发生变化，需要重新计算阈值。首先，使用更新后的模型计算控制集样本的异常分数。然后，通过校准器得到新的校准分数集合。最后，重新计算 q 分位数作为新的阈值。动态阈值机制确保了检测标准与环境同步：当正常行为演化导致分数分布整体偏移时，阈值随之调整，将新的正常行为纳入正常范围，降低误报；而真正的攻击由于其分数显著高于阈值，仍能被有效检测。

3 实验与评估

3.1 数据集与实验设置

(1) 实验数据集

实验采用俄克拉荷马大学发布的以太坊网络流量数据集^[35-36]，该数据集由包含从 10 个全球观测点收集的 3 个月以太坊网络流量数据，时间跨度为 2024 年 8 月至 2024 年 11 月。数据包括网络层会话记录、节点指标以及蜜罐交互日志。每条记录包含时间戳、源 IP、目标 IP、协议类型、数据包数、字节数、节点 ID 等字段。数据预处理步骤如下：

1) 划分时间窗口。将流量数据按时间维度进行划分，共得到 90 个时间窗口。选择前 7 个时间窗口数据用于框架的初始训练期（仅使用正常样本），后续共 83 个窗口，每天作为独立的时间点，对应文中标记的测试时间点 T1 至 T83。

2) 会话聚合。以（源 IP，目标 IP，协议）三元组作为区块链网络会话标识，对每个会话按时间排序，提取所需 24 维特征，每个会话形成一个特征向量序列。序列长度固定为 $T=10$ ，不足长度则补零，超过则截断。

3) 特征标准化。使用训练集计算每个特征的均值和标准差，对全部数据进行标准化处理。

4) 异常标注。利用蜜罐日志，将蜜罐捕获的恶意 IP 相关的会话标记为攻击。同时，为了模拟正态性偏移，在第 30 个时间窗口和第 60 个时间窗口引入模拟的“协议升级”事件，通过调整部分特征的分布，使协议升级后新协议类型的出现反映在特征分布中，但不改变异常标签。

5) 模拟协议升级。为模拟区块链网络真实协议升级带来的正常行为分布偏移，分别在第 30、

第 60 时间窗口触发两次模拟协议升级事件, 仅修改正常会话样本的特征分布, 所有样本的异常/攻击标签保持不变, 保证模拟偏移仅来源于正常的业务演化, 不混入攻击行为。

□ 选择调整的特征集合: 结合以太坊协议升级 (P2P 通信版本迭代、区块广播消息格式更新) 对网络流量的影响规律, 选取 7 个受协议版本迭代影响的流量、协议维度特征进行分布扰动, 分别为 Packets_per_second (每秒数据包数)、Bytes_per_packet (平均数据包大小)、Byte_rate_variance (字节速率方差)、Packet_size_entropy (数据包大小分布熵)、Is_tcp、Is_udp、Protocol_entropy (协议分布熵), 节点状态类特征 (CPU、内存、对等节点数量等) 保持原始分布不变, 模拟协议升级不会直接改变节点硬件运行状态的现实情况。

□ 原始分布与模拟扰动参数设置: 原始训练窗口 (T1-T29) 中, 上述 7 个特征均服从训练集拟合得到的多元高斯分布 $\mathcal{N}(\mu_{old}, \Sigma_{old})$ 。在 T30、T60 时间窗口及之后窗口, 对上述 7 个特征进行分布平移扰动:

● 均值向量: $\mu_{new} = 1.4 \times \mu_{old}$, 对特征均值放大 1.4 倍;

● 协方差矩阵: $\Sigma_{new} = 1.2 \times \Sigma_{old}$, 适度放大方差, 模拟协议升级后会话流量的波动程度小幅提升;

● 约束: 所有特征扰动后取值不违背物理含义 (数据包数、字节数等非负; 熵值约束在理论取值区间), 若计算得到非法数值, 则截断至合法取值边界。

模拟仅对正常会话样本执行上述分布重采样扰动。攻击样本全部维持原始特征不变, 攻击标签完全不变, 确保实验中分布偏移是正常业务演化导致, 而非攻击带来的变化。

□ 模拟设置与真实区块链协议演化关系

以太坊等区块链网络进行协议升级 (例如 P2P 网络通信协议迭代、新的区块广播/交易传播消息类型上线) 时, 节点间会话报文长度、报文发送速率、TCP/UDP 协议占比、数据包大小的混乱度会发生系统性改变。但是, 节点的 CPU、内存占用、对等节点连接数量这类节点硬件状态指标不会发生突变。本次模拟通过修改流量与协议相关特征的均值与协方差, 复现协议升级后正常流量统计分布整

体发生平移、波动增大的现象。同时保留攻击样本原有特征与标签, 用来检验算法能否区分 “正常协议演化造成分布偏移” 和 “真实攻击造成异常”, 与论文所要解决的真实业务场景完全对齐。

最终实验选取的数据集包含有约 80 万条区块链网络会话记录。其中, 攻击样本约 5 万条, 其余样本为正常样本。攻击类型包括日食攻击 (约 1.5 万条)、DDoS 攻击 (约 2 万条)、交易泛洪攻击 (约 1 万条)、其他恶意行为 (约 0.5 万条)。表 5 总结了数据集的基本统计信息。

表 5	数据集描述	
	统计项	数值
	总会话数	802,347
	正常样本	752,134
	攻击样本	50,213
	日食攻击	15,231
	DDoS 攻击	20,112
	交易泛洪攻击	10,045
	其他攻击	4,825
	时间窗口	90
	平均每天会话数	8,915
	特征维度	24

(2) 实验设置

BADT-SSR 框架的实验在 Intel Xeon Gold 5218 CPU 和 NVIDIA V100 GPU 上进行测试, 依托 PyTorch 1.10 实现。框架的超参数分别设置为隐藏层维度 $H=64$, 序列长度 $T=10$, dropout 率 0.2, 校准器分箱数 $K=20$, 置换检验次数 $N_{perm}=1000$, EWC 正则化强度 $\lambda=1000$, 微调轮数 5 轮, 微调学习率 10^{-4} , 动态阈值取 95% 分位数。

为验证 BADT-SSR 框架的有效性, 本文从传统静态异常检测、不同范式持续学习异常检测、通用流式概念漂移检测、原始特征直接分布检验四个维度选取对比基线。所有方法使用相同的训练集进行初始训练, 并在后续每个时间点使用相同数量样本, 实验重复 5 次取平均值, 基线方法说明如下:

1) Static-AE: 使用初始训练的时序自编码器, 不进行任何更新, 直接应用于所有时间窗口。代表静态模型, 模型训练完成后不再适配后续流量变化, 用于量化动态自适应机制相对静态模型的整体

性能提升。

2) **Retrain**: 每个时间点使用该时间点的全部正常样本重新训练模型。代表全量重训练策略, 每窗口从零重新训练; 用于对比“完整重训练”与本文受控EWC更新在性能、开销间的差异。

3) **Fine-tune**: 每个时间点用当前窗口的正常样本微调模型, 不进行EWC处理。代表无正则微调策略, 无持续学习约束, 用于验证EWC正则对于缓解灾难性遗忘的作用。

4) **EWC-AE**: 采用EWC正则化的自适应模型, 但无偏移检测模块, 强制每个时间点更新。经典正则化持续学习算法, 采用强制逐窗口更新, 不经过偏移检测闸门; 用来验证偏移触发受控更新相比无条件强制更新的增益。

5) **OCSVM**: 支持向量机, 使用BADT-SSR框架相同特征, 不考虑时序。

6) **Isolation Forest**: 孤立森林, 基于特征空间的随机分割, 同样使用窗口平均特征。

7) **LwF-AE**: 持续学习基线, 基于时序自编码器架构, 采用LwF持续学习策略。不配置偏移检测模块, 每个时间窗口强制执行模型微调, 利用旧模型蒸馏输出软标签约束新模型, 抑制灾难性遗忘, 代表经典知识蒸馏型持续学习异常检测方法。

8) **MAS-AE**: 持续学习基线, 时序自编码器基础上引入MAS持续学习正则, 通过样本梯度度量参数重要性, 对重要参数施加惩罚; 无偏移检测模块, 每个窗口强制更新, 代表梯度重要性权重约束类持续学习异常检测算法。

9) **ADWIN-AE**: 流式概念漂移检测基线, ADWIN为经典自适应滑动窗口流式漂移检测算法。将时序自编码器输出的原始异常分数作为单变量流式输入, 采用ADWIN检测异常分数序列的概念漂移。当检测漂移触发模型更新, 采用普通微调更新AE模型, 代表主流单变量流式漂移检测结合自编码器的方案。

10) **DDM-AE**: 流式概念漂移检测基线, DDM基于分类错误率监控实现流式概念漂移检测。将每个时间窗口异常样本识别错误率作为监控指标, 触发漂移后微调时序自编码器, 是数据流漂移检测经典基线。

11) **KS-Baseline**: 直接分布检验异常检测基线, 不使用自编码器, 直接对原始特征做窗口级

KS两样本检验, 对比当前窗口特征分布与历史控制集分布, 将KS统计量作为异常分数完成异常判定; 代表直接基于原始特征分布检验的异常检测方案。

12) **KL-Raw-Baseline**: 直接分布检验异常检测基线, 无自编码器、无时序嵌入, 直接对原始特征做分箱, 计算当前窗口与历史控制集之间KL散度作为异常得分, 完成异常识别; 为直接原始特征KL分布检验基线。

(3) 评估指标

1) **AUC (Area Under ROC Curve)**: 随机给定一个正样本和一个负样本, 用分类器分类和预测, 该正样本得分比负样本得分大的概率。它是衡量模型区分样本为正常和攻击的能力, 不受阈值选择影响, 取值范围[0,1], 越高越好。

2) **F1-score**: 精确率和召回率的调和平均, 综合反映检测准确性, 计算公式为:

$$F1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

3) **误报数 (FP)**: 正常样本被误判为攻击的数量, 反映误报水平。

4) **漏报数 (FN)**: 攻击样本被漏判的数量, 反映漏报水平。

5) **偏移检测准确率**: 对于已知偏移事件, 检测模块是否能正确发出偏移信号 ($p < 0.05$)。

6) **平均检测延迟**: 从攻击发生到首次检测到的时间窗口差值, 反映响应的及时性。

3.2 实验结果与分析

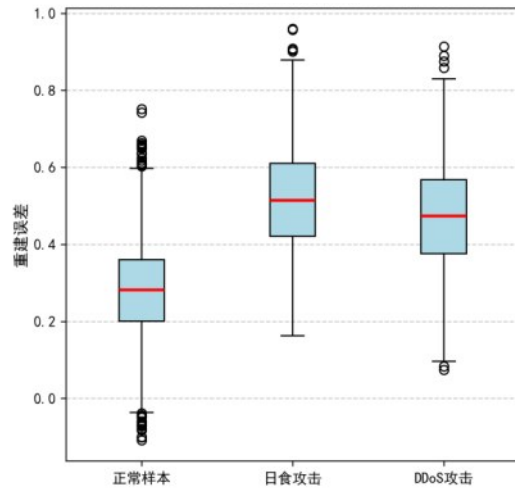
(1) 样本空间构建性能分析

样本空间构建的质量直接决定了后续偏移检测和攻击判定的上限。一个理想的样本空间应满足两个条件: 一是正常样本的重建误差尽可能小且集中, 表明模型准确编码了正常行为的时序模式; 二是异常样本的重建误差尽可能大, 与正常样本形成明显分离, 确保异常可被有效检出。本节通过对比时序自编码器与全连接自编码器在上述两方面的表现, 验证时序建模对样本空间质量的提升作用。

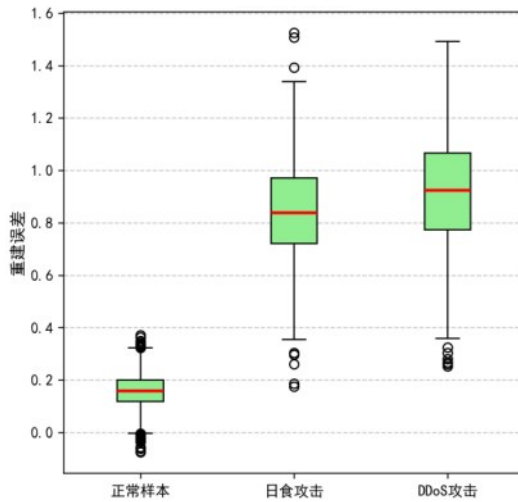
实验中使用训练集前7个时间窗口数据训练两种自编码器, 训练完成后不再更新。测试集采用第8个至第90个时间窗口的全部数据。为直观展示样本空间质量, 从测试集中随机抽取5000条正常样本、500条日食攻击样本和500条DDoS攻击样本,

计算每条样本的重建误差, 绘制箱线图对比两类模型的误差分布。

图2展示了两种自编码器在不同类型样本上的重建误差分布情况。在全连接自编码器中, 正常样本的误差分布比较分散, 中位数为0.28, 四分位距为0.16。日食攻击样本误差中位数为0.52, DDoS攻击样本误差中位数为0.48, 日食攻击、DDoS攻击样本误差与正常样本均存在一定重叠, 区分边界较为模糊。在时序自编码器中, 正常样本的误差显著降低且更集中, 中位数为0.16, 四分位距为0.08, 而日食攻击样本误差中位数为0.85, DDoS攻击样本误差中位数为0.92, 日食攻击和DDoS攻击样本误差明显偏高, 与正常样本形成了清晰的分离。



(a) 全连接自编码器



(b) 时序自编码器

图2 两类模型的误差分布对比图

为进一步量化分离程度, 我们定义分离度指标 $Sep = (\mu_a - \mu_n) / \sqrt{\sigma_n^2 + \sigma_a^2}$, 其中, μ_n 、 σ_n^2 为正常样本误差的均值和方差, μ_a 、 σ_a^2 为攻击样本误差的均值和方差。该指标值越大, 表示正常与异常的区分度越好。全连接自编码器分离度平均为1.2, 而时序自编码器分离度平均提升至2.8。

上述结果表明, 时序自编码器通过捕捉交易广播的突发性、区块传播的周期性等区块链特有的时序特征, 使正常样本的编码更紧凑, 异常样本的偏离更显著。相比忽略时序的全连接自编码器, 时序自编码器构建的样本空间具有更高的表示质量, 为后续的偏移检测和攻击判定奠定了更可靠的基础。

(2) 概率校准性能分析

为验证保序回归校准方法的有效性, 实验分别从分布可比性和偏移检测能力两个维度对无校准、Sigmoid校准和保序回归校准进行对比。在初始训练期的前7个时间窗口、第一次协议升级后的第30个时间窗口和第二次协议升级后的第60个时间窗口分别选取正常样本。在应用三种不同的校准方式后, 采用KS检验比较各窗口的分数分布是否来自同一分布。在分布可比性方面, 图3(a)展示了三种校准方式下第7个时间窗口分别与第30、第60个时间窗口正常样本分数分布的KS检验p值。KS检验的原假设为两组样本来自同一分布, p值越大表明分布越一致。无校准方式下, T7 vs T30和T7 vs T60的p值分别为0.003和0.001, 境外远小于0.05, 说明协议升级后正常样本的异常分数分布与训练期存在显著差异, 分布可比性丧失。Sigmoid校准后, 两组对比的p值分别上升至0.031和0.028, 虽有一定的改善但仍低于0.05, 分布差异依然较为显著, 表明了Sigmoid函数对长尾分布的拟合能力不足。而保序回归校准后, 两组对比的p值分别达到0.124和0.109, 均大于0.05, 各窗口分数分布可比性显著提升, 无法拒绝分布一致的原假设。这表明保序回归校准有效消除了协议升级带来的分数分布漂移, 使不同时期的异常得分具备了可比性。

需要特别说明的是本文实验中采用的KS检验与KL散度结合置换检验两套统计推断方案的实验目的、比较对象存在本质区别, 二者并不矛盾, 分别服务于概率校准效果评估与分布偏移检测两个不同环节。

1) KS 检验（用于评价概率校准效果）

目的：评估保序回归是否消除由正常行为演化造成的异常分数系统性整体漂移，提升不同时间窗口样本分数的分布可比性；

比较对象：协议升级前后均为正常样本的校准分数集合（T7 正常样本 vs T30 正常样本、T7 正常样本 vs T60 正常样本）；

判定含义：KS 检验原假设两组样本来自同一分布。保序回归校准后 $p > 0.05$ ，不能拒绝原假设，代表正常样本经过校准后，消除了协议升级带来的分数整体平移，不同时间窗口的正常样本分数可以放在同一尺度下对比。需要注意的是该环节参与对比的全部是正常样本，KS 检验只用来验证校准是否抹平正常行为演化引发的系统性分数偏移，而不是用来检测异常事件。

2) KL 散度+置换检验（用于样本空间偏移检测）

目的：在校准完成、分数具备可比性基础上，识别当前窗口整体样本（允许混杂正常新行为与攻击样本）相对于历史控制集的真实分布形态、概率密度结构的显著差异；

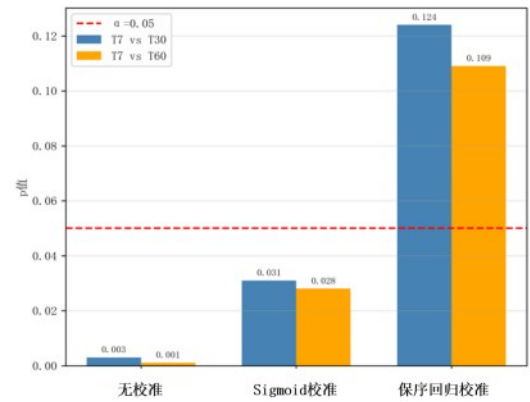
比较对象：历史控制集（历史正常样本校准分数） vs 当前时间窗口完整样本集（包含当前全部会话样本，既包含新模式正常样本，也可能混入攻击样本）；

判定含义：置换检验原假设两组样本来自同一分布。当 $p < 0.05$ 拒绝原假设，判定发生显著样本空间偏移。

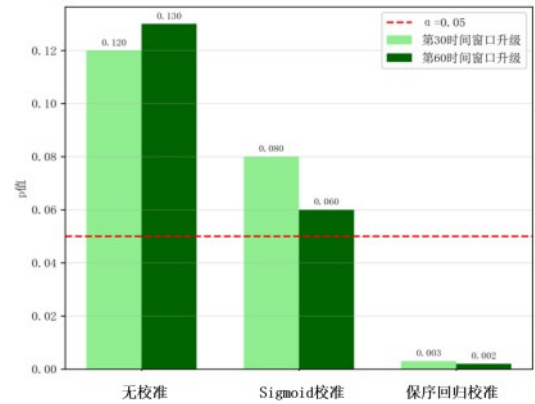
保序回归概率校准完成的是分数尺度归一化处理，将不同时间窗口正常样本的异常分数映射至统一的[0,1]评价标尺，消除因协议升级等正常事件带来的分数整体平移，也就是提升分布可比性（对应 KS 检验 p 值上升）；但校准并不会抹除样本内部真实的分布结构差异。在统一标尺之上，KL 散度+置换检验继续去度量两组集合概率分布形状的差异。

在协议升级场景下，当前窗口包含大量遵从新协议模式的正常样本，即便每个样本的分数被映射至统一标尺，样本分数的直方图、概率密度形态依然会和历史控制集产生结构性差别。因此在 KS 检验验证“正常样本跨窗口分数可比”的前提下，KL-置换检验依然可以检出协议升级带来的真实分

布结构偏移。简言之，KS 检验验证“尺子是同一把”，而 KL 散度+置换检验用这把统一的尺子，去衡量两组数据分布形状到底有没有真正不同。若不经保序回归校准，原始异常分数本身标尺随时间漂移，KL 散度计算会被虚假的整体分数平移干扰，无法区分“标尺漂移”和“真实分布结构变化”，这也正是概率校准作为偏移检测前置步骤的必要性。



(a) KS 检验 p 值



(b) 偏移检验 p 值

图3 分布可比性与偏移检测能力对比

偏移检测能力方面，偏移检测模块采用 KL 散度量当前窗口与控制集的分布差异，并通过置换检验评估该差异的统计显著性。KL 散度越大则表明分布差异越大，置换检验将观测的 KL 散度与随机排列下的 KL 散度进行比较，输出 p 值作为显著性判据。图3 (b) 展示了第30和第60个时间窗口协议升级后三种校准方式下的偏移检测 p 值。无校准方式下，虽然 KL 散度较大，但由于原始分数缺乏概率校准， p 值分别为0.12和0.13，未能达到 $\alpha =$

0.05 的显著水平,说明直接使用原始分数进行偏移检测并不可靠。Sigmoid 校准下, p 值分别达到 0.08 和 0.06, 同样未通过显著性检验, 原因在于 Sigmoid 校准对分布偏移的校正不够充分, 部分掩盖了真实变化的幅度。保序回归校准则能够成功检测出偏移, 两个升级窗口的 p 值分别为 0.003 和 0.002, 均远低于 0.05。这是因为保序回归校准消除了正常分布漂移对分数的影响后, KL 散度能够更准确地反映当前窗口与控制集之间的真实分布差异, 置换检验因此能够识别出统计上显著的偏移信号。上述结果表明, 保序回归校准不仅使异常分数具备分布可比性, 更为基于 KL 散度与置换检验的偏移检测提供了可靠输入, 是连接样本空间构建与自适应更新的关键环节。

图 4 直观展示了三种校准方式下三个时间窗口的分数分布对比。T15 为协议升级前的正常窗口, T30 和 T60 分别为两次模拟协议升级事件所对应的时间窗口。在无校准条件下, T30 的分数分布从均值 0.048 变为 0.175, 相较 T15 出现显著右移, 峰值位置和分布形态均存在明显差异, 表明原始重建误差分数受协议升级影响发生了系统性偏移, 不同窗口的分数不具备可比性。Sigmoid 校准虽将分数压

缩至了固定区间, 但 T30 与 T15、T60 之间的分布依然存在较大差异, 实验结果显示 T30 均值 0.236, T60 均值 0.119, T15 均值 0.091, 说明固定形状的 S 型函数无法灵活适应不同窗口的分数分布变化, 未能有效消除协议升级带来的分布漂移。保序回归校准后, T30 与 T60 的分数分布接近, 均值分别为 0.388 和 0.396, 中位数分别为 0.462 和 0.388, 分布形态和数值范围趋于一致。T15 的分布均值是 0.286 亦与上述两窗口高度接近, 仅因自身为升级前的真实正常数据而略有下移。这一现象的根本原因在于保序回归不预设函数形式, 能够直接根据校准窗口的真实异常标签拟合出灵活的单调映射曲线, 从而有效校正了因正常行为演化导致的分布漂移。上述结果表明, 保序回归校准能够使不同时间窗口的分数具有可比性, 为后续基于校准分数的偏移检测提供了可靠基础。

图 5 绘制了保序回归的校准函数曲线。蓝色散点表示训练集中每个正常样本的原始异常分数及其对应的经验误报率 FPR, 红色阶梯线为保序回归拟合得到的单调递减校准函数, 满足原始异常分数越高, 校准输出分数越低的单调性特点。图中标注示例点中原始重建异常分数为 0.50, 经过保序回归映

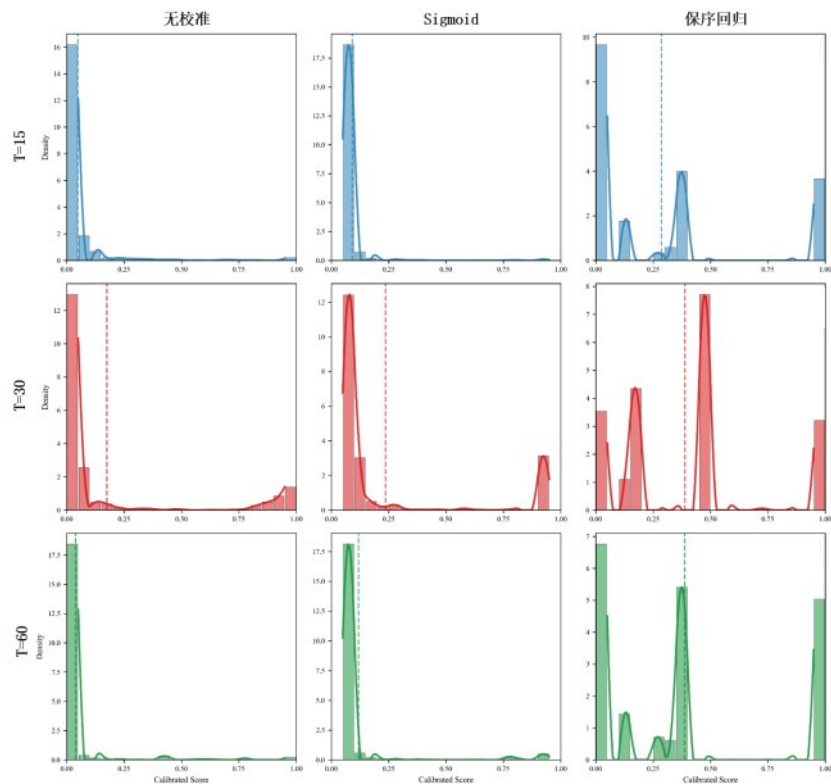


图4 不同时间窗口的分数分布对比

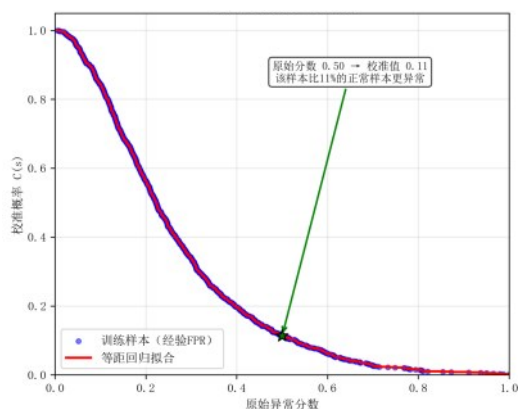


图5 保序回归的校准函数曲线

射得到校准分数0.11，代表以原始分数0.50作为检测阈值，训练集仅11%的正常样本会被误判为异常，即该样本偏离程度高于89%的历史正常样本，异常嫌疑更高。

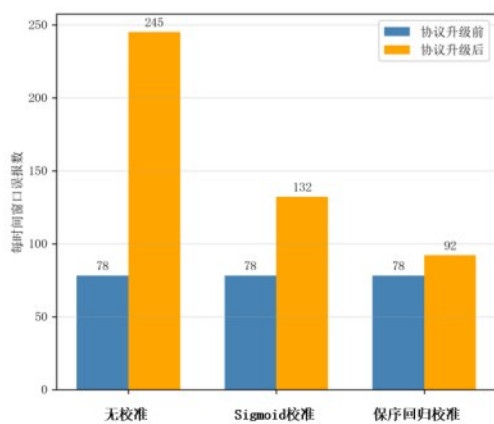


图6 不同校准方式对攻击检测性能的影响

图6为协议升级前后的每日误报数变化，进一步评估了不同的校准方式对攻击检测性能的影响情况。在误报率方面，无校准方式协议升级后误报率激增，每个时间窗口的误报数量从78上升到了245。Sigmoid校准下误报率有所缓解，每日误报从78上升至132。而保序回归校准下的误报率基本保持稳定，每日误报从78上升至92。保序回归校准与其他两种校准方式的误报率差距在协议升级后进一步拉大，这说明通过有效的概率校准并能有效降低正常偏移导致的误报率。

(3) 偏移检测性能分析

偏移检测模块在90个时间窗口中共触发15次偏移告警，其中2次对应模拟协议升级，其余13次为流量自然波动导致的误告警。偏移检测的主要作

用是提醒系统关注潜在变化，而非直接决策。系统仅需对少量告警进行确认，即可触发必要的模型更新。

表6 消融实验(平均AUC)

配置	AUC (均值±标准差)	95% 置信区间
BADT-SSR 完整	0.89±0.012	[0.875, 0.905]
无校准器	0.82±0.017	[0.800, 0.840]
无偏移检测	0.84±0.014	[0.823, 0.857]
无EWC	0.79±0.020	[0.767, 0.813]
无LSTM	0.81±0.016	[0.791, 0.829]

我们通过消融实验验证各模块的贡献，表6总结了消融实验结果。无校准器为直接使用原始异常分数进行偏移检测和阈值设定，AUC降至0.82。无偏移检测为强制每个时间点更新，AUC为0.84，但更新次数从2次增至90次，模型计算开销大幅增加。无EWC的AUC降至0.79，说明EWC对防止模型遗忘至关重要。无LSTM（改用全连接AE模型）的AUC降至0.81，验证了时序建模的必要性。

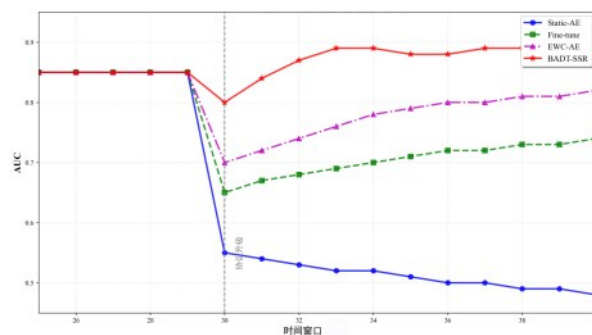


图7 协议升级事件前后AUC变化曲线

由偏移检测性能实验可知，本文偏移检测模块对预设协议升级真实偏移事件实现了召回，但区块链网络天然存在流量随机波动，实验90个时间窗口中共产生13次由流量自然波动引发的虚假告警，偏移检测精确率仅为13.3%。正如前文所述，偏移检测模块定位为变化告警提示单元，不直接输出模型更新决策，偏移告警触发后必须经过告警确认环节才允许启动样本空间自适应更新，避免随机流量波动引发不必要的模型更新。本节结合区块链网络运维实际场景，阐述告警确认完整流程、告警过滤

策略及抑制无效更新的实现方案。

(1) 偏移告警的多级确认流程

当置换检验输出 $p < 0.05$ 触发偏移告警后, 执行由“自动初筛、链上事件匹配、人工复核”组成的三级确认流程, 只有全部判定为“正常业务演化偏移”, 才允许调用自适应更新模块, 判定为随机波动、攻击行为, 则拒绝模型更新。

1) 自动初筛。首先校验本次偏移告警对应的统计指标, 包括 KL 散度观测值、置换检 p 值、当前窗口样本数量。设置两个工程阈值最小 KL 散度阈值、窗口最小样本量。若观测 KL 散度小于最小 KL 散度阈值, 或者当前时间窗口会话样本数低于窗口最小样本量, 直接判定为流量随机波动, 丢弃本次告警, 不进入后续确认流程。样本量过低会造成直方图分箱统计不稳定, 是产生虚假告警的重要诱因。在本实验数据集上, 经调优设置最小 KL 散度阈值为 0.12, 窗口最小样本量为 2000 时, 可过滤大部分由小样本随机抖动带来的虚假告警。

2) 链上事件自动匹配。经过自动初筛保留的告警, 进一步对接区块链链上事件日志, 自动检索告警时间窗口附近是否发生协议升级、硬分叉、新 DApp 大规模上线、节点版本批量更新等业务层面正常演化事件。系统维护事件时间索引, 比对告警时间窗口与链上事件时间戳的时间差, 若时间差小于预设时间容忍窗口 (本系统设置为 2 个时间窗口), 则标记为疑似正常业务偏移。若检索到大量恶意 IP 接入、攻击事件告警, 则标记为疑似攻击导致分布变化, 直接拒绝模型更新, 将该窗口标记为攻击监测重点窗口, 只输出安全告警, 不更新样本空间。

3) 运维人员人工复核。当自动匹配无法得到明确结论, 无匹配链上事件, KL 散度又超过阈值, 将告警连同该窗口校准分数直方图、KL 散度、置换检验 p 值、链上原始日志摘要推送给区块链网络运维人员做人工复核。运维人员结合链外运维日志, 判断分布偏移来源是偶然流量波动、新型攻击, 还是未录入事件库的正常业务变更。只有人工确认属于正常业务演化, 才触发自适应模型更新。

(2) 抑制不必要模型更新的配套机制

除多级告警确认流程外, 框架还配套多重约束机制, 进一步降低不必要模型更新频次:

1) 更新冷却机制: 完成一次自适应模型更新

之后, 设置更新冷却周期, 冷却周期内即便再次触发偏移告警, 也不会执行模型更新, 仅记录告警日志。本文实验设置冷却周期为 5 个时间窗口, 避免短时间内连续多次更新模型, 抑制由短暂流量震荡引发的连续更新。

2) 控制集容量与滑动窗口机制: 控制集保存历史正常样本校准分数, 设置最大存储的容量上限, 当样本超出上限采用滑动窗口丢弃最旧的数据样本。既保留历史正常行为模式, 又可以缓慢适配缓慢漂移, 避免微小波动触发频繁更新。

3) 更新触发双重条件约束: 模型自适应更新需要同时满足两个条件: 一是置换检验检测到统计显著偏移。二是经过三级确认流程判定偏移来源于正常业务演化。仅偏移告警不足以触发更新, 从逻辑层面杜绝流量波动直接驱动模型更新。

(3) 实验数据集下的效果分析

将上述工程策略回用到本文 90 个时间窗口实验数据集进行后验分析, 原始实验共计 15 次偏移告警 (2 次真实协议升级、13 次自然流量波动虚假告警)。引入自动初筛之后, 13 次虚假告警中 11 次可以在自动初筛阶段直接过滤, 剩余 2 次虚假告警进入链上事件匹配环节, 因无对应链上业务事件而被拦截。2 次真实协议升级告警均可以完整保留, 不会被过滤。经过整套确认过滤流程之后, 最终仅触发 2 次模型更新, 与真实的业务偏移事件完全对齐, 完全消除由自然流量波动带来的不必要模型更新。同时需要说明, 该套策略属于部署层面的工程补充, 不改变论文核心算法, 核心算法的实验验证结果保持不变、不受影响。

(4) 自适应更新性能分析

图 7 展示了第 30 时间窗口模拟协议升级前后各方法的 AUC 变化曲线。其中, 横轴为时间窗口, 纵轴为 AUC。Static-AE 在升级后 AUC 从 0.85 骤降至 0.55, 误报激增。Fine-tune 适应后略有回升但无法恢复至原有水平。EWC-AE 缓慢回升, 但需多次更新。BADT-SSR 在检测到偏移后, 立即触发 EWC 微调, AUC 在 2 个时间窗口内回升至 0.88, 并保持稳定。第 60 个时间窗口的第二次升级时呈现类似趋势, 验证了方法的鲁棒性。

(5) 攻击检测性能分析

表 7 展示了完整的 BADT-SSR 对各类攻击的检测 AUC、召回率和精确率。其中, 日食攻击检测

效果最好, AUC 达 0.94, 因其流量模式与正常流量模式差异明显。交易泛洪攻击较难检测, AUC 为 0.86, 因其可能模仿正常高频交易。

表7 不同攻击类型的检测性能

攻击类型	AUC	召回率	精确率
日食攻击	0.94	0.91	0.88
DDoS 攻击	0.91	0.89	0.85
交易泛洪攻击	0.86	0.82	0.79
其他攻击	0.82	0.78	0.75

具体案例中, 第45天节点A的会话特征出现异常: `packets_per_second` 达120 (正常值20-50), `src_ip_diversity` 仅2 (正常 ≥ 15), `bytes_per_packet` 仅64字节 (正常800-1500), `peer_churn_rate` 达0.8 (正常0.1)。输入时序自编码器, 重建误差为0.85, 校准后分数0.97, 超过动态阈值0.92, 判定为攻击。经核实, 该节点遭受日食攻击, 攻击者仅使用两个IP发送大量空包。BADT-SSR成功捕获该攻击。在第62天, 节点B在短时间内收到大量TCP SYN包: `packets_per_second` 达500 (正常30), `in_connections` 达300 (正常10), `src_ip_diversity` 为50但每个IP只发少量包, `bytes_per_packet` 仅40字节, `connection_success_rate` 仅0.2 (正常0.9)。重建误差0.92, 校准后0.99, 判定为攻击。经分析为SYN洪水DDoS攻击。第30天, 以太坊协议升级导致区块传播机制变化, 节点间出现新型消息, 特征变化如下: `total_bytes` 略有增加, `protocol_entropy` 从0.8升至0.95 (新协议类型出现), `peer_churn_rate` 短暂升高。偏移检测p值0.003, 触发告警。分析师确认为正常升级后, BADT-SSR利用当日正常样本微调模型, 更新后误报恢复正常。图8详细的展示了BADT-SSR对不同攻击类型的检测性能。

为了验证能否有效捕捉区块链流量正常模式, 本实验根据设计好的基线进行对比。表8展示了各方法在整个90天时间窗口上的平均AUC和F1-score, 以及平均每日误报/漏报数。其中, ★表示经双样本t-检验, 本文BADT-SSR与该基线相比性能提升具有统计显著性。从实验结果可以看出, BADT-SSR在AUC和F1上均优于所有基线方法, 平均AUC达到0.89, 比静态模型Static-AE提升约

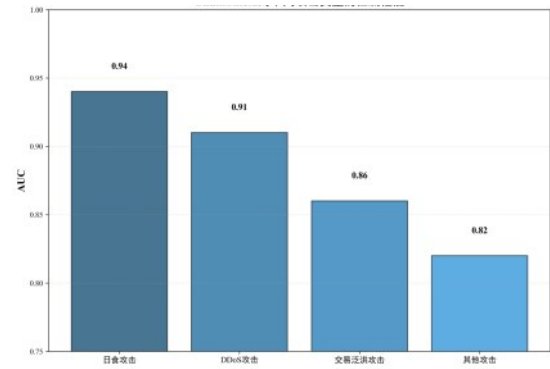


图8 对不同攻击类型的检测性能图

44%, 比无自适应的OCSVM和Isolation Forest提升25%以上。

BADT-SSR的每日误报数量和漏报数量是最少的。其中, FP=51、FN=9, 表明其在保持高检测率的同时, 有效的控制了误报规模。Retrain性能虽接近于BADT-SSR, 其AUC为0.86, 但需要在每个时间点对全部正常样本进行标注, 该方法的标注成本极高。Fine-tune由于无正则化, 在适应新分布时发生灾难性遗忘现象, 检测性能劣于EWC-AE和BADT-SSR。EWC-AE性能略低于BADT-SSR。这是因为EWC-AE本身是每天更新, 但缺少偏移检测, 所以性能略低, 说明偏移检测模块避免了不必要的频繁更新, 减少了性能波动和计算开销。LwF-AE与MAS-AE两类持续学习异常检测方法, 依靠蒸馏/参数权重约束缓解灾难性遗忘, 但缺少偏移检测前置判断逻辑, 每个时间窗口强制执行模型更新。在区块链网络无真实偏移的时间窗口也会执行微调, 引入流量噪声污染历史正常模式, 平均AUC仅分别为0.77、0.78, FP也显著高于BADT-SSR。ADWIN-AE、DDM-AE是流式概念漂移检测方案, 将自编码器输出异常分数作为一维数据流监控漂移。ADWIN-AE取得AUC=0.80, 优于DDM-AE (AUC=0.76)。但两类流式漂移检测面向单变量数据流设计, 区块链流量为多维混合时序特征, 正常流量随机波动容易触发虚假漂移告警, 引发不必要模型更新, 性能仍然显著低于BADT-SSR (AUC=0.89)。KS-Baseline、KL-Raw-Baseline属于直接分布检验异常检测方法, 不使用时序自编码器做特征嵌入, 直接基于原始维特征做分布对比。原始区块链网络流量噪声高、维度高, 直接分布检验难以区分协议升级带来的正常分布偏移与网络攻击, 平均AUC仅0.72、0.73, 协议升级场景下误报

激增。相比之下, BADT-SSR 框架采用时序自编码器完成高维时序特征嵌入, 保序回归实现分数标尺对齐, KL-置换偏移检测模块仅在检测统计显著偏移之后才触发 EWC 持续学习更新, 避免无意义频繁模型更新, 兼顾抗遗忘与漂移识别能力, 因此取得最优检测性能。

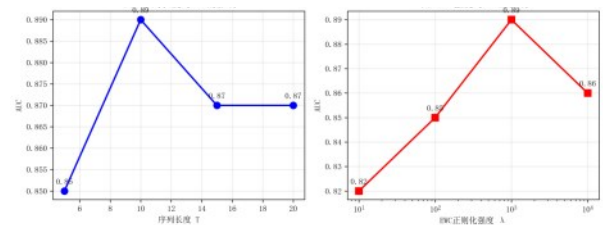
(6) 参数敏感性与计算开销分析

为了更好分析不同超参数对 BADT-SSR 框架的影响情况, 对关键超参数进行敏感性分析。我们分别测试了序列长度 T 、EWC 正则化强度 λ 、校准分箱数 K 及动态阈值分位数 q 对模型性能影响。

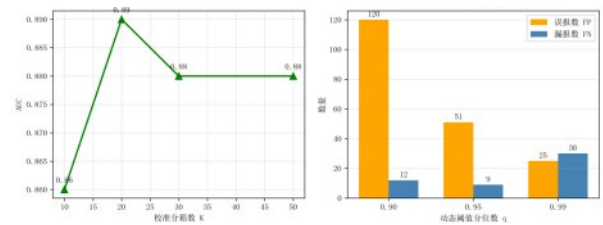
序列长度 T 决定了检测模型可追溯的时序范围。当 T 取 5 时, 模型因时序信息不足, AUC 仅为 0.85。当 T 增至 10, 模型能充分捕捉交易广播的突发性和区块传播的周期性, AUC 达到最高值 0.89。继续增大 T 至 20, 过长的序列引入噪声, AUC 略微回落至 0.87。因此, 选择 T 等于 10。EWC 正则化强度 λ 平衡了模型对旧知识的保留与新分布的适应。当 λ 等于 10 时, 正则化过弱, 模型在协议升级后产生严重遗忘, AUC 降至 0.82。当 λ 等于 1000 时, 效果最佳, AUC 为 0.89。当 λ 等于 10000 时, 正则化过强, 模型适应新分布缓慢, AUC 仅为 0.86。故取 λ 等于 1000。校准分箱数 K 将影响 KL 散度计算时分布近似的精细度。实验中当 K 等于 10 时, 分布刻画过于粗糙, AUC 为 0.86。当 K 等

于 20 时, 分布可更准确刻画分布之间的差异, AUC 达到 0.89。当 K 等于 50 时, 易受到训练样本的噪声干扰, AUC 则略降至 0.88。因此, 设定 K 等于 20。动态阈值分位数 q 决定了理论误报率上界。当 q 等于 0.95 时, 在误报数 FP 和漏报数 FN 均取得良好平衡, 分别为 51 和 9。当 q 等于 0.9 时, 误报增加至 120。当 q 等于 0.99 时, 漏报上升至 25。故选择 q 等于 0.95。图 9 直观展示了上述四个参数对 AUC、误报、漏报的影响。

在计算开销方面, 各模块平均运行时间为特征



(a) 序列长度对 AUC 影响 (b) EWC 强度对 AUC 影响



(a) 分箱数对 AUC 影响 (b) 分位数对误报/漏报影响

图 9 关键超参数对模型性能的影响

表 8

各方法检测性能对比

方法	AUC (均值±标准差)	95% 置信区间	F1 (均值±标准差)	平均 FP/时间窗口	平均 FN/时间窗口	显著性	年份
Static-AE	0.62±0.021	[0.596,0.644]	0.58±0.023	245	31	★	2020
Retrain	0.86±0.015	[0.843,0.877]	0.81±0.016	78	14	★	2022
Fine-tune	0.79±0.018	[0.769,0.811]	0.74±0.019	132	22	★	2023
EWC-AE	0.84±0.014	[0.824,0.856]	0.79±0.015	92	18	★	2023
OCSVM	0.71±0.022	[0.685,0.735]	0.65±0.024	187	28	★	2023
Isolation Forest	0.68±0.024	[0.653,0.707]	0.62±0.025	210	30	★	2023
LwF-AE	0.77±0.016	[0.751,0.789]	0.74±0.017	184	23	★	2024
MAS-AE	0.78±0.015	[0.763,0.797]	0.75±0.016	176	22	★	2022
ADWIN-AE	0.80±0.014	[0.784,0.816]	0.77±0.015	161	20	★	2024
DDM-AE	0.76±0.017	[0.739,0.781]	0.73±0.018	192	24	★	2024
KS-Baseline	0.72±0.020	[0.697,0.743]	0.69±0.021	233	31	★	2025
KL-Raw-Baseline	0.73±0.019	[0.708,0.752]	0.70±0.020	224	30	★	2025
BADT-SSR	0.89±0.012	[0.875,0.905]	0.85±0.014	51	9	—	—

提取时间开销 0.5 秒、异常分数计算时间开销 1.2 秒、分数校准时间开销 0.1 秒、偏移检测时间开销 0.8 秒、模型微调时间开销约 30 秒。90 个时间窗口中仅触发 2 次微调，每次自适应更新的额外开销约 20 s，两次更新合计额外更新开销约 40 s，满足准实时监测需求。而 EWC-AE 强制每天更新，更新总时间达 2700 秒。计算开销结果如图 10 所示。

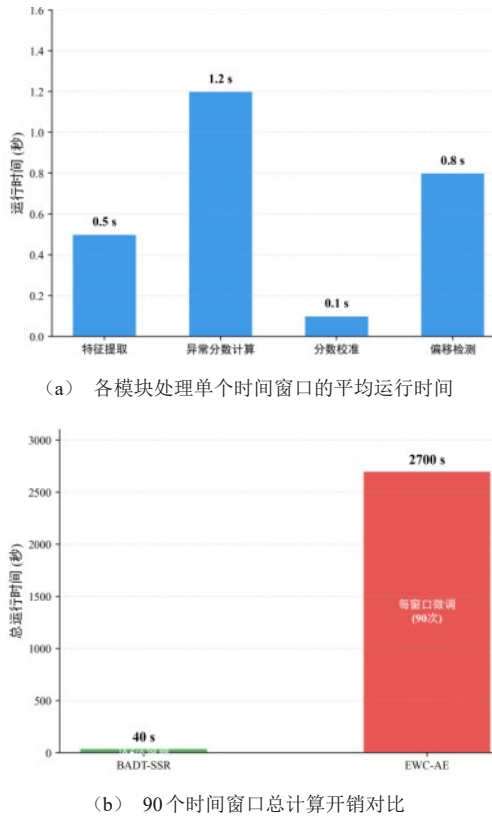


图 10 计算开销分析

3.3 局限性分析

(1) 蜜罐标签噪声影响分析

本文数据集正常样本采用“未被蜜罐 IP 标记”作为判定准则。蜜罐采集机制存在一定局限，部分攻击行为若不与蜜罐 IP 产生交互，将造成攻击样本漏捕，漏捕攻击样本会被误标记为正常样本，混入训练集与控制集，引入标签噪声。该标签噪声会对各模块产生不同程度影响：1) 样本空间构建：训练集中混入少量漏捕攻击样本，会使得重建误差基准被抬高，扩大正常样本重建误差的离散程度；2) 保序回归概率校准：噪声样本会改变经验 FPR 分布，轻微改变校准分数映射关系；3) KL-置换偏移检测：控制集混入异常样本会改变基准分布，

会提升虚假偏移告警概率；4) 动态阈值：控制集噪声样本会抬高分位数阈值，会带来漏报风险。

本框架通过两类机制缓解标签噪声危害：1) 训练前做简单异常预过滤：剔除重建误差显著离群的样本，降低训练集噪声占比。2) 更新触发采用多重严格约束，即使控制集存在噪声，单一统计偏移告警无法触发 EWC 更新，避免噪声进一步扩散进入后续迭代的模型与控制集。3) 受公开数据集条件限制，无法获取完全无噪声的真实标签。未来研究可引入多源情报融合，进一步降低标签噪声。

(2) 攻击者伪造正常流量诱导更新风险分析

攻击者若刻意生成统计特征高度酷似新业务正常模式的恶意流量，试图制造样本空间分布偏移，希望触发 EWC 自适应更新，将恶意模式吸收进模型与控制集，后续同类攻击将面临难以检出的风险。从本文给出的形式化触发条件可以看出，仅依靠偏移检测输出 $Flag_{shift}=True$ 无法触发模型更新，攻击者想要诱导模型更新，必须同时满足如下约束：1) 自动初筛约束：攻击者制造的偏移需要通过偏移幅度、样本规模等自动初筛过滤条件，小规模、短时爆发的伪造偏移会被自动初筛直接拦截；2) 链上事件佐证约束：窗口必须存在协议升级、新应用上线等真实链上业务事件。如果没有对应的真实链上事件记录，即使统计偏移告警存在，也无法满足触发条件；攻击者很难伪造真实链上业务事件记录；3) 人工复核约束：需要运维人员人工复核确认该偏移属于正常业务演化。

4 总结

本文针对区块链网络异常检测中攻击手段持续演化、网络行为动态偏移、标注数据样本稀缺等挑战，提出了一种基于样本空间重构的区块链异常流量检测技术，通过样本空间构建、概率校准、偏移检测、自适应更新构建完整的区块链异常检测流程。首先，通过时序自编码器对历史正常流量进行无监督学习，将以太坊交易广播等典型区块链网络行为的时序特征编码为可度量的重建误差，构建反映正常行为的样本空间表示。其次，引入保序回归将原始异常分数映射为具有明确概率意义的校准值，解决不同时间窗口检测结果不可比的问题，并采用 KL 散度、置换检验进行统计推断，准确识别由协议升级、新应用上线等区块链特有因素导致的

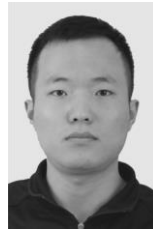
样本空间偏移,有效区分正常演化与真正攻击。最后,引入弹性权重固化机制,在适应新样本空间的同时保留原有正常模式记忆,避免遗忘旧的有价值检测模式。实验结果验证了本文所提出方法的有效性,异常检测的平均AUC达到0.89,较静态模型提升约40%,误报率降低70%以上,且在协议升级等正常变化发生后能够快速恢复检测性能。未来工作将围绕交易层、网络层、合约层等多源区块数据融合和自动化偏移检测机制结合,进一步提升方法的实用性和鲁棒性。

参考文献:

- [1] Mazumder P T. Blockchain in trade finance: reducing fraud and improving efficiency through digital ledger technology[J]. Digital Finance, 2025, 7(4): 1043-1063.
- [2] J A, Isravel D P, Sagayam K M, et al. Blockchain for healthcare systems: Architecture, security challenges, trends and future directions[J]. Journal of Network and Computer Applications, 2023, 215: 103633.
- [3] Bouachir O, Aloqaily M, Ö Ö, et al. FederatedGrids: Federated Learning and Blockchain-Assisted P2P Energy Sharing[J]. IEEE Transactions on Green Communications and Networking, 2022, 6(1): 424-436.
- [4] Nandanwar H, Katarya R. Optimized intrusion detection and secure data management in IoT networks using GAO-Xgboost and ECC-integrated blockchain framework[J]. Knowledge and Information Systems, 2025, 67(10): 9531-9586.
- [5] 刘敖迪, 杜学绘, 王娜等. 区块链系统安全防护技术研究进展[J]. 计算机学报, 2024, 47(03): 608-646.
- [6] Xu C, Jin G, Lu R, et al. A federated learning architecture for blockchain DDoS attacks detection[J]. IEEE Transactions on Services Computing, 2024, 17(5): 1911-1923.
- [7] 高昊昱, 曹春杰, 白伊瑞等. 区块链安全监管研究综述[J]. 通信学报, 2025, 46(04): 49-70.
- [8] Rehman Z, Gregory M A, Gondal I, et al. A Private Ethereum Testbed for Network Traffic Dataset Collection[C]. 2025 IEEE 35th International Telecommunication Networks and Applications Conference (ITNAC), 2025: 1-8.
- [9] He D, Gong W, Tu C, et al. HASHL: Dynamic Hash Verification for Detecting and Preventing Eclipse Attacks[J]. IEEE Internet of Things Journal, 2025, 12(13): 23524-23535.
- [10] Gai K, Zhang G, Jiang P, et al. A Moving Target DDoS Defense Approach in Consortium Blockchain[J]. IEEE Transactions on Dependable and Secure Computing, 2025, 22(4): 4362-4378.
- [11] Khoa T V, Son D H, Hoang D T, et al. Collaborative Learning for Cyberattack Detection in Blockchain Networks[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024, 54(7): 3920-3933.
- [12] Mounnan O, Manad O, Boubchir L, et al. A review on deep anomaly detection in blockchain[J]. Blockchain: Research and Applications, 2024, 5(4): 100227.
- [13] Hassan M U, Rehmani M H, Chen J. Anomaly Detection in Blockchain Networks: A Comprehensive Survey[J]. IEEE Communications Surveys & Tutorials, 2023, 25(1): 289-318.
- [14] Chithanuru V, Ramaiah M. An anomaly detection on blockchain infrastructure using artificial intelligence techniques: Challenges and future directions - A review[J]. Concurrency and Computation: Practice and Experience, 2023, 35(22): e7724.
- [15] 戴千一, 张斌, 郭松等. 基于多分类器集成的区块链网络层异常流量检测方法[J]. 通信学报, 2023, 44(03): 66-80.
- [16] Su J, Chen Z, Zhu K, et al. Anomaly detection for blockchain nodes based on eBPF and fine-tuning large language model[J]. Journal of Information Security and Applications, 2026, 97: 104329.
- [17] Li M, Chen J, Guo S, et al. DDoS Attack Detection in Business Blockchain Networks: A Review, Framework, and Challenges[J]. IEEE Transactions on Network Science and Engineering, 2026, 13: 906-929.
- [18] Liu C, Xu Y, Sun Z. Directed dynamic attribute graph anomaly detection based on evolved graph attention for blockchain[J]. Knowledge and Information Systems, 2024, 66(2): 989-1010.
- [19] Chen S, Liu Y, Zhang Q, et al. Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions[J]. Advanced Intelligent Systems, 2025, 7(8): 2400898.
- [20] Du H, Shen M, Liu Y, et al. Fine-Grained and Class-Incremental Malicious Account Detection in Ethereum via Dynamic Graph Learning[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 10130-10145.
- [21] Li Z, Zhuo S, Shi Y, et al. Unveiling Blockchain Transactions Insights: Behavioral Anomaly Detection via Relational Mechanisms[J]. IEEE Internet of Things Journal, 2025, 12(24): 52706-52719.
- [22] Hu X, Shu Z, Tong Z, et al. Fine-grained Ethereum behavior identification via encrypted traffic analysis with serialized backward inference [J]. Computer Networks, 2023, 237: 110110.
- [23] Kim J, Nakashima M, Fan W, et al. A Machine Learning Approach to Anomaly Detection Based on Traffic Monitoring for Secure Blockchain Networking[J]. IEEE Transactions on Network and Service Management, 2022, 19(3): 3619-3632.
- [24] 朱龙隆, 陈翔, 陈浩东等. 基于近似测量算法 Sketch 的高精度区块链异常检测机制[J]. 计算机研究与发展, 2024, 61(10): 2526-2539.
- [25] Voronov T, Raz D, Rottenstreich O. A Framework for Anomaly Detection in Blockchain Networks With Sketches[J]. IEEE/ACM Transactions on Networking, 2024, 32(1): 686-698.
- [26] Shen X, Xu C, Zhu L. Blockchain Anomaly Transaction Detection Method Based on Graph Continual Learning[J]. IEEE Transactions on Network Science and Engineering, 2026, 13: 6059-6078.
- [27] Wang R, Zhang Y, Peng L. Anomaly Detection Service for Blockchain Transactions Using Minimal Substitution-Based Label Propagation[J]. IEEE Transactions on Services Computing, 2024, 17(5): 2054-2066.
- [28] Liu J, Wu J, Chen J, et al. Trans2Graph: Mining Ethereum Phishers With Graph on Heterogeneous Temporal Transaction Data[J]. IEEE Transactions on Information Forensics and Security, 2025, 20: 12950-12963.
- [29] Zhang M, Sun W, Li H, et al. Manipulated Transaction Collision Attack on Execute-Order-Validate Blockchain[J]. IEEE Transactions on Dependable and Secure Computing, 2025, 22(2): 1132-1149.
- [30] 郭虎升, 张旭, 张羽桐等. 多类型概念漂移引导的在线分时适应性集成[J]. 计算机研究与发展, 2026, 63(4): 987-997.
- [31] 祁晓博, 陈佳明, 史颖等. 基于主动-被动增量集成的概念漂移适应方法[J]. 自动化学报, 2025, 51(5): 1131-1144.

- [32] Cabrera Martin I, Mukherjee S, Baimagambetov A, et al. Evolving machine learning in non-stationary environments: A unified survey of drift, forgetting, and adaptation[J]. Applied Artificial Intelligence, 2026, 40(1): 2707047.
- [33] Wang L, Zhang X, Su H, et al. A comprehensive survey of continual learning: Theory, method and application[J]. IEEE transactions on pattern analysis and machine intelligence, 2024, 46(8): 5362-5383.
- [34] Wu B, Ding Z, Huang J. A review of continual learning in edge AI[J]. IEEE Transactions on Network Science and Engineering, 2026.
- [35] Seidenberger S, Maiti A. EtherBee: A Global Dataset of Ethereum Node Performance Measurements Coupled with HoneyPot Interactions and Full Network Sessions[C]. 2025 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), 2025: 1-3.
- [36] Seidenberger S, Maiti A. Initial Evidence of Pervasive Reconnaissance Targeting Ethereum Node Infrastructure[C]. Proceedings of the 7th ACM International Symposium on Blockchain and Secure Critical Infrastructure, 2025: Article 3.

[作者简介]



刘敖迪（1992- ），男，博士，信息工程大学讲师，主要研究方向为区块链安全、数据安全。

于佳鑫（2004- ），男，辽宁长春人，信息工程大学本科、主要研究方向为区块链安全。

杜学绘（1968- ），女，河南辉县人，博士，信息工程大学教授、博士生导师，主要研究方向为网络信息安全。

王娜（1980- ），女，河南济源人，博士，信息工程大学副教授、硕士生导师，主要研究方向为网络信息安全。

曹利峰（1981- ），男，河南禹州人，博士，信息工程大学教授、博士生导师，主要研究方向为网络信息安全。