文章类型：论文|刊号（ISSN）：2972-4236(P) / 2972-4244(O)

人工智能驱动的网络数据泄露主动检测与溯源技术研究

陈泓羽
中国联合网络通信有限公司青岛市分公司
DOI:10.32629/acair.v4i1.19349

[摘　要] 网络数据泄露风险的常态化对检测与溯源技术提出精准化、智能化、隐私化的刚性需求。本文构建多维度协同的主动检测技术体系,通过特征提取融合、算法创新优化与隐私保护架构设计,实现泄露行为的精准识别。突破区块链存证、知识图谱追踪与多模态分析的溯源技术瓶颈,建立全链路追溯机制。创新动态基线自适应、隐私增强与可解释性增强技术,提升检测溯源的适应性、安全性与可信度。研究成果为网络数据安全防护提供技术支撑,具有重要的学术价值与实践意义。
[关键词] 人工智能；网络数据泄露；主动检测；溯源技术
中图分类号：TP18　文献标识码：A

Research on AI-driven Proactive Detection and TracingTechnology for Network Data Breaches

Hongyu Chen
China United Network Communications Co., Ltd. Qingdao Branch

[Abstract] The increasing prevalence of network data breach risks necessitates more precise, intelligent, and privacy-preserving detection and tracing technologies. This paper constructs a multi-dimensional collaborative proactive detection technology system, achieving accurate identification of breach behaviors through feature extraction and fusion, algorithm innovation and optimization, and privacy-preserving architecture design. It overcomes the bottlenecks of blockchain-based evidence storage, knowledge graph tracking, and multimodal analysis in tracing technologies, establishing a full-link traceability mechanism. Innovative dynamic baseline adaptation, privacy enhancement, and interpretability enhancement technologies improve the adaptability, security, and credibility of detection and tracing. The research results provide technical support for network data security protection and have significant academic and practical value.
[Key words] Artificial intelligence; Cyber data breach; Proactive detection; Source tracing technology

引言

数字化转型的纵深推进使数据成为核心生产要素,网络数据的开放性与流动性加剧泄露风险。数据泄露呈现攻击手段隐蔽化、传播路径复杂化、影响范围扩大化的特征,传统依赖规则匹配的被动防御模式难以应对动态变化的安全威胁。主动检测与溯源技术作为数据安全防护的核心环节,其性能直接决定风险防控的有效性。现有检测技术存在特征提取片面、算法适应性不足、跨域协同能力薄弱等问题,溯源过程面临数据可信度不足、关联分析困难、隐私保护缺失等挑战。人工智能(Artificial Intelligence,AI)技术凭借强大的特征学习、模式识别与自适应能力,为突破上述瓶颈提供可能。基于此,本文整合多领域技术成果,构建AI驱动的主动检测与溯源技术体系,通过多维度特征融合、创新算法设计、隐私增强架构与全链路溯源机制,实现网络数据泄露的精准检测、快速溯源与有效防控,为网络空间数据安全保障提供全新技术路径。

1 检测技术体系构建

1.1多维度特征提取与融合

终端行为特征的全面捕捉是精准检测的基础,卷积神经网络-循环神经网络(Convolutional Neural Network-Recurrent Neural Network,CNN-RNN)混合模型兼具空间特征提取与时序依赖分析能力,可深度挖掘终端进程的系统调用序列、文件操作轨迹与网络连接模式,实现时空特征的协同提取。自编码器(Autoencoder,AE)通过无监督学习方式对高维特征进行降维重构,剔除冗余信息,保留具有判别性的核心特征,提升模型训练效率与检测精度。

用户行为的动态性要求检测模型具备基线自适应能力,长短期记忆网络(Long Short-Term Memory,LSTM)模型通过学习用户长期行为数据,构建涵盖访问时段、操作频率、数据交互范围的个性化行为基线。注意力机制可聚焦偏离基线的关键行为片段,实现异常行为的精准定位,解决传统检测中异常特征淹没于

正常数据的问题。某电子商务平台应用该技术,通过分析用户登录地理位置、设备类型等多维度信息,成功识别多起异地异常登录事件,有效防范账户盗用引发的数据泄露。

加密流量的隐蔽性给检测带来挑战,针对这一问题,提取流量的统计特征、时序特征与会话特征构建多维度融合模型。统计特征包含数据包大小分布、传输频率等量化指标,时序特征捕捉流量的时间依赖关系,会话特征聚焦通信交互的上下文信息[1]。通过特征级融合策略,实现加密流量中敏感数据传输行为的有效识别,破解传统技术难以穿透加密层的困境。某电力企业采用该方法,成功检测出通过加密隧道传输敏感运行数据的异常行为,保障了电力网络数据安全。

1.2异常检测算法创新

Web流量是数据泄露的主要传播路径,基于Transformer的检测框架采用自注意力机制,可并行处理超文本传输协议/安全超文本传输协议(HyperText Transfer Protocol/HyperText Transfer Protocol Secure,HTTP/HTTPS)流量的统一资源定位符(Uniform Resource Locator,URL)参数编码、请求头字段等应用层信息,构建细粒度流量基线。结合业务场景特征优化模型决策逻辑,强化对结构化查询语言(Structured Query Language,SQL)注入、文件包含等漏洞利用行为的识别能力,实现高级持续性威胁(Advanced Persistent Threat,APT)攻击等高级威胁的精准检测。该框架通过分析攻击链关键环节的交互特征,有效捕捉低频率、隐蔽性强的攻击行为,弥补传统检测对未知威胁响应滞后的不足。

数据泄露的传播渠道呈现多元化特征,单一渠道检测难以形成防控合力。集成学习方法通过融合邮件、即时通信、网盘等多渠道的深度学习检测模型,构建跨渠道协同检测体系。各基础模型针对特定渠道的传输特征进行优化训练,集成模块通过加权投票机制整合检测结果,提升复杂场景下的检测准确性。某大型企业应用该体系,实现对员工通过多种通信工具外发敏感商业数据行为的全面监测,检测覆盖度较单一模型提升显著[2]。

网络环境的动态变化要求检测策略具备自适应调整能力,强化学习技术通过构建环境－动作－奖励的交互机制,动态优化检测规则与参数配置。当网络拓扑、传输协议或攻击手段发生变化时,模型通过持续学习调整检测阈值与特征权重,维持检测性能的稳定性。在某云计算平台的应用中,该技术根据租户访问模式的变化实时调整检测策略,有效适应多租户共享环境下的复杂安全态势。

1.3联邦学习隐私保护

跨域数据共享与隐私保护的矛盾制约着检测技术的规模化应用,基于联邦学习(Federated Learning,FL)的分布式检测架构采用“本地训练－参数聚合”模式,实现数据不出域前提下的模型协同训练。各参与节点利用本地数据训练模型参数,仅上传模型更新信息至聚合服务器,通过加密算法保障参数传输安全,从源头规避数据泄露风险。该架构在金融行业的应用中,实现了多家银行在不共享客户数据的情况下,联合构建信用卡欺诈检测模型,有效提升了跨机构数据泄露检测能力[3]。

差分隐私(Differential Privacy,DP)技术通过在模型参数传输过程中添加可控噪声,实现隐私保护与模型性能的动态平衡。噪声强度根据隐私保护等级动态调整,确保攻击者无法通过参数反推原始数据,同时保证聚合后的模型仍具备良好的检测性能。某医疗数据平台应用该技术,在联合训练患者数据泄露检测模型时,成功保护了患者个人健康信息的隐私安全,模型检测准确率未出现显著下降。

模型更新过程的安全性是FL架构的关键,构建多级安全验证机制保障训练过程可信。采用数字签名技术验证参与节点的身份合法性,通过区块链记录参数更新日志实现全程可追溯,利用同态加密(Homomorphic Encryption,HE)技术保障参数聚合过程的安全性。该机制有效防范恶意节点篡改模型参数、伪造检测结果等攻击行为,确保分布式检测体系的稳健运行。

2 溯源技术突破

2.1区块链溯源存证

数据全生命周期的可追溯性是溯源的核心要求,基于区块链(Blockchain,BC)的溯源存证系统采用分布式账本技术,记录数据采集、存储、传输、使用等各环节的操作日志。区块通过密码学哈希算法串联,形成不可篡改的链式结构,确保溯源数据的完整性与真实性。某电子商务平台应用该系统,完整记录用户敏感信息的访问日志,在数据泄露事件发生后,通过账本回溯快速定位访问源头,为责任认定提供可靠依据[4]。

智能合约(Smart Contract,SC)技术为溯源过程提供自动化执行保障,预先设定溯源验证规则与流程,当满足触发条件时自动执行身份验证、权限核查等操作。合约内置的逻辑判断机制确保溯源过程的透明可信,避免人为干预导致的溯源结果失真。在供应链数据溯源中,SC自动验证各环节数据传输的合法性,一旦发现异常传输行为立即触发预警并记录相关证据,提升溯源响应效率。

多链协同场景下的溯源需求日益增长,跨链溯源协议通过互操作接口实现不同BC网络间的数据互通。采用哈希锁定、公证人机制等技术,建立跨链数据的映射关系,支持多源溯源数据的协同验证与路径追踪。某跨区域物流企业应用该协议,实现了不同区域BC节点中货物信息传输记录的统一溯源,成功追踪到多起物流信息泄露的传播路径。

2.2知识图谱智能追踪

实体关系图谱是实现关联溯源的核心载体,基于图神经网络(Graph Neural Network,GNN)技术构建涵盖用户、设备、数据资源等实体的关系网络,精准刻画实体间的访问关系、传输路径等关联信息。通过图嵌入技术将实体与关系映射到低维向量空间,强化对隐蔽关联的挖掘能力,实现异常行为关联实体的智能追踪。某网络安全企业利用该技术,构建了黑客攻击团伙的关联图谱,成功追踪到多起数据泄露事件的幕后操控者。

社区发现算法通过挖掘关系图谱中的密集连接子图,识别

隐蔽的异常行为关联网络。该算法基于模块度优化原则,将具有协同行为的实体聚类形成社区结构,揭示分散在不同网络节点中的关联攻击行为。在某政务数据平台的溯源应用中,通过该算法发现多个看似无关的账户存在隐性关联,成功揪出利用多个账户协同窃取政务敏感数据的犯罪团伙[5]。

图算法的应用提升了溯源的效率与精度,最短路径算法可快速定位数据泄露的源头节点与传播路径,风险节点预测算法通过分析实体的关联强度、行为频率等特征,识别潜在的风险传播节点。某金融机构结合这两种算法,在数据泄露事件发生后,仅用数小时就完成了从泄露点到源头账户的全路径追溯,并提前预警了多个可能被波及的关联账户,最大限度地降低了泄露影响。

2.3多模态溯源分析

数据泄露的溯源涉及文本、图像、日志等多源异构数据,多模态溯源分析模型通过跨模态特征对齐算法,实现不同类型数据的关联分析。采用注意力机制强化模态间的语义关联,将文本数据的语义特征、图像数据的视觉特征、日志数据的结构化特征映射到统一特征空间,构建多维度溯源证据链。某互联网企业应用该模型,通过整合系统日志、通信记录、文件内容等多源数据,成功还原了敏感用户信息的泄露全过程[6]。

跨模态数据的异构性给关联分析带来挑战,为此开发特征对齐与融合算法。在特征对齐阶段,通过模态转换、语义映射等技术消除不同数据类型的结构差异。在融合阶段,采用加权求和、注意力融合等策略整合多模态特征,强化溯源关键信息的表征能力。某医疗影像平台采用该技术,实现了对泄露医疗影像数据的传输路径追踪,通过关联分析影像文件的元数据、传输日志等多源信息,精准定位泄露源头。

时空特征分析模块为溯源提供精准的定位与时序依据,空间特征包含数据传输的网络地址、设备位置等地理信息,时间特征捕捉行为发生的时序关系。通过时空联合分析,构建数据迁移的时空路径图谱,实现泄露行为的全程可视化追踪。在某交通数据平台的应用中,该模块通过分析数据传输的时间序列与地理位置变化,成功追踪到敏感交通流量数据的泄露路径,为后续防控措施的制定提供了精准支撑。

3 关键技术创新

3.1动态基线自适应

传统静态基线难以适应网络行为的动态变化,基于强化学习的动态基线调整机制通过持续学习用户与系统的行为模式演变,实现基线的实时更新与优化。模型将行为变化视为环境状态转移,通过奖励机制引导基线向贴合实际行为的方向调整,解决静态基线易导致误报、漏报的问题。某电商平台应用该机制,根据用户购物高峰期的行为模式动态调整检测基线,在保障检测效果的同时降低了误报率。

检测精度与误报率的平衡是技术应用的关键,自适应阈值调整算法通过分析历史检测数据的分布特征,动态优化决策阈值。当网络安全态势趋于严峻时,适当降低阈值提升检测灵敏度。当正常行为模式发生波动时,合理提高阈值减少误报干扰。该算法在某能源企业的应用中,成功平衡了电力监控系统数据泄露检测的准确性与系统运行的稳定性。

在线学习模块为模型持续进化提供支撑,通过增量学习技术处理新增数据,无需重新训练即可实现模型参数的更新。采用滑动窗口机制筛选有效训练数据,保留近期关键行为特征,剔除过时信息对模型的干扰。某通信企业应用该模块,使检测模型能够持续吸收新的攻击行为特征,实现对新型数据泄露手段的快速响应。

3.2隐私增强技术

HE技术实现加密数据上的模型训练与推理,无需解密即可对加密状态的敏感数据进行处理,从根本上保障数据在检测过程中的隐私安全。采用部分同态加密或全同态加密方案,根据检测任务的计算复杂度选择适配的加密算法,在加密强度与计算效率之间寻求平衡。某科研机构应用该技术,在联合分析多家医院的医疗数据泄露风险时,实现了患者隐私信息的全程加密保护。

安全多方计算协议支持跨机构协同溯源分析,多个参与方在不泄露本地数据的前提下,通过加密计算实现溯源信息的联合处理。采用秘密共享、混淆电路等技术构建计算框架,确保各参与方仅能获取自身权限范围内的溯源结果,无法推导出其他机构的敏感数据。某跨区域金融监管场景中,该协议实现了多家银行在隐私保护前提下的联合溯源,成功追踪到跨机构数据泄露的传播路径。

零知识证明机制为溯源验证提供隐私保护支持,证明方在不泄露溯源关键信息的前提下,向验证方证明溯源结果的真实性。采用非交互式零知识证明方案,简化验证流程,提升溯源结果的验证效率。在某政务数据溯源系统中,该机制确保了敏感数据的溯源过程不泄露政务信息的核心内容,同时保障了溯源结果的可信度。

3.3可解释性增强

模型的黑箱特性制约着检测结果的采信度,基于沙普利可加解释(SHapley Additive exPlanations,SHAP)值的解释模块通过量化特征对检测结果的贡献度,实现检测结果的可视化解释[7]。采用热力图、特征重要性排序等可视化方式,直观呈现导致异常判定的关键特征与决策逻辑,帮助安全人员理解模型判断依据。某网络安全公司应用该模块,为检测到的异常数据传输行为提供详细的特征贡献度分析,显著提升了安全人员的研判效率。

决策路径追踪算法实现异常行为的全流程追溯,记录模型从特征提取、特征融合到决策输出的完整过程。通过还原模型的推理链条,定位导致检测结果产生的关键环节,为检测结果的复核与模型优化提供支撑。在某大型企业的应用中,该算法成功追溯到模型误判的根源是某类正常业务流量的特征与异常行为相似,为模型参数调整提供了精准依据。

人机交互界面为安全人员深度参与分析过程提供支撑,设

计简洁直观的操作界面，支持安全人员对检测结果进行标注、修正与反馈。界面集成特征可视化、决策路径展示等功能，实现人机协同的检测与溯源分析。某安全运营中心应用该界面，使安全人员能够快速介入异常行为分析，通过人工经验补充模型不足，提升整体防控效果。

4 结论

本文构建了AI驱动的网络数据泄露主动检测与溯源技术体系，通过多维度特征提取融合、异常检测算法创新与FL隐私保护架构，实现了泄露行为的精准识别。突破BC溯源存证、GNN智能追踪与多模态溯源分析技术，建立了全链路、可信化的溯源机制。创新动态基线自适应、隐私增强与可解释性增强技术，提升了技术体系的适应性、安全性与可信度。该技术体系整合了电子商务、电力、金融等多个领域的实践经验，通过理论创新与技术落地的深度融合，为网络数据安全防护提供了全新解决方案。在实际应用中，成功实现了对多种场景下数据泄露行为的有效检测与快速溯源，验证了技术的实用性与可靠性。未来研究可聚焦三个方向：一是强化模型在边缘计算环境下的部署能力，提升终端侧数据泄露检测的实时性。二是探索生成式AI在未知威胁检测中的应用，增强对新型泄露手段的识别能力。三是深化跨领域技术融合，推动检测溯源技术与零信任架构、数字孪生等技术的协同发展，构建更为全面的网络数据安全防护体系。

[参考文献]

[1]赵雪娇,李浩升,马怡璇.基于模糊RBF神经网络的网络数据泄露检测[J].中国新技术新产品,2025(3):140-142.

[2]袁国泉.基于自适应深度学习网络的电网攻击数据安全检测[J].印刷与数字媒体技术研究,2025,(S2):129-140+176.

[3]张媛.计算机存储数据泄露漏洞自动检测方法[J].数字技术与应用,2025,43(08):66-68.

[4]郭舒扬.基于改进PSO-PFCM聚类算法的大数据平台隐私泄露检测方法[J].自动化与仪器仪表,2025,(05):269-272.

[5]刘建莉.大数据时代人工智能技术在电子商务网络安全防范中的应用研究[J].电子元器件与信息技术,2025,9(4):251-253.

[6]李伟周.人工智能驱动的数据泄露检测技术在网络安全中的应用研究[J].网络安全和信息化,2025,(02):18-20.

[7]鲁凯,钮艳,宋增人.人工智能数据安全检测技术研究[J].保密科学技术,2025,(02):5-11.

作者简介：

陈泓羽(1990--),女,汉族,山东青岛人,硕士研究生,工程师,研究方向：系统信息工程学专业方向。