

DOI:10.19344/j.cnki.issn1671-5276.2025.05.034

AI 技术在网络空间安全防御中的应用研究

孔德丽

(南京机电职业技术学院, 江苏 南京 211306)

摘要:人工智能技术在网络空间安全保障方面面临物联网漏洞、AI 模型隐患及复杂智能攻击等挑战。AI 在入侵检测、安全分析等方面展现出强大潜力。提出应对措施包括:建立 AI 驱动的监控预警平台,即时防御威胁;引入高级身份验证与智能化访问控制,加固网络安全体系。

关键词:人工智能;网络空间安全;防御技术

中图分类号:TP29 **文献标志码:**B **文章编号:**1671-5276(2025)05-0173-04

Research on Application of AI Technology in Cyberspace Security Defense

KONG Deli

(Nanjing Vocational Institute of Mechatronic Technology, Nanjing 211306, China)

Abstract:With regards to the challenges such as IoT vulnerabilities, AI modeling pitfalls and complex intelligence attacks that AI technology in cyberspace security are facing, and the AI's strong potentialities in intrusion detection, security analysis and other aspects, this paper proposes the countermeasures including the establishment of an AI-driven monitoring and warning platform for immediate defense against threats, and the introduction of advanced authentication and intelligent access control to strengthen the network security system.

Keywords:artificial intelligence; cyberspace security; defense technology

1 人工智能时代网络空间安全面临的风险

1.1 攻击技术的升级—AI 驱动下的新型网络安全威胁

AI 革新了网络攻击,使恶意软件的危害更智能、隐蔽,其会自动探测漏洞,高效绕过防护,自我学习进化,定制化攻击提升成功率^[1],这些都对网络安全构成了严峻挑战。

1.2 AI 模型安全隐患—数据污染与算法安全挑战

AI 安全面临数据污染与对抗样本威胁,前者植入误导信息^[2],后者诱使模型误判^[3],以此凸显 AI 脆弱性与决策风险,亟需产学研合力,深化研究,构筑安全防线。

1.3 关键设施智能化带来的网络空间安全风险剧增

工业互联网与智慧城市发展,加剧了关键设施网络安全风险。物联网设备多点易攻^[4],危及系统稳定,会造成巨大经济损失,动摇社会与国家安全。须研创智能防护^[5],立标规范,筑智能 AI 安全墙,应战威胁,确保设施安全。

1.4 物联网与智能设备漏洞—扩大化的网络攻击入口

随着物联网扩张使网络安全^[6]复杂化,弱设备成为黑客靶心,智能弊病引来独立攻击、DDoS 风暴及网络渗透。需要 AI 实时监控分析,强固件更新,加密通信,建全防护,提升意识,共抗新威胁。

2 基于人工智能的网络空间安全防御关键技术

2.1 机器学习驱动的入侵检测系统

由机器学习驱动的入侵检测系统 (ML-based IDS) 是 AI 在网络安全领域的关键应用^[7]。此类系统持续收集网络流量数据 (如包头信息、协议行为),经数据预处理 (清洗、转换、整合) 以供高质量的机器学习模型训练^[8]。模型学习数据模式,能有效区分正常与异常行为,识别 0 Day 攻击、木马等威胁。该系统能够实时推断、分类,适应新攻击策略,并自我完善,强化网络防御能力。

2.2 深度学习的威胁情报分析

深度神经网络 (DNNs) 从海量数据中学习、提

作者简介:孔德丽(1982—),男,江苏泰州人,副研究员,硕士,研究方向为高校信息化规划、建设与管理,17044358@qq.com。

取关键特征,揭示隐蔽攻击手段^[9],可应用于威胁检测时,深度学习挖掘网络流量、日志和 API 调用中的深层模式,识别潜在威胁,提升智能防火墙与动态防御策略的效果。基于深度学习的零信任系统能够适应环境变化,学习新型攻击,实时更新防御,快速响应未知威胁,加速威胁识别与应对。

2.3 自动化响应与防御策略优化

采用自动化与智能化响应技术实时监测网络环境,一旦检测到威胁或预测到安全事件,立即触发预设自动化程序,有效处置风险^[10]。借助智能决策支持,系统根据预设条件明智决策,包括优选业务连续性与灾难恢复策略,既加快了事件响应速度,又减少了人工干预,大幅提升了网络安全防御的灵活性与时效性。

2.4 自然语言处理(NLP)在安全事件关联分析中的应用

异常检测与行为分析技术运用机器学习与大数据实时监测、深度剖析网络活动,识别异常模式以追踪潜在攻击^[11]。系统构建数据行为模型与正常基准,全方位监控网络状态,异常偏离则触发警报并应对,确保网络安全稳定。结合 UEBA,深入理解用户/实体常规行为模式^[12],对比分析变化,无明显攻击特征下亦能揭示权限滥用、数据泄漏等异常,提升事件识别准确率与响应速度,有力增强防护系统的综合性能,构筑网络空间的坚实防线。

2.5 预测性维护与漏洞管理

预测性维护与智能漏洞管理是 AI 网络安全防御核心技术,其运用机器学习与大数据监测、预测网络设备与系统状态,防患于未然。预测性维护通过分析故障模式、性能数据,预测故障窗口、隐患威胁,提前规划设备更新维护^[13],降低老化漏洞引发的风险。智能漏洞管理则全面挖掘、评估、修复系统漏洞,自动定期/不定期扫描识别,按严重程度、重要性排序优先级,制定更新策略^[14]。系统通过持续学习、优化模型,准确预测未来漏洞,提前制定应对措施。

2.6 智能蜜罐与欺骗防御技术

智能蜜罐是网络安全主动防御的重要手段,通过模拟真实网络资源和服务,诱骗并捕获非法攻击者。相较传统蜜罐^[15],智能蜜罐融入 AI 和机器学习技术,具有更强的真实感和交互性,能动态模拟各种系统环境,迷惑攻击者。获取的攻击

数据用于研究攻击手段、追溯黑客行为、预测未来威胁^[16]。欺骗防御技术扩展蜜罐概念,创建虚拟资产和假目标,可误导攻击者,确保核心资产安全;即使防线被突破,损失的只是蜜罐^[17]。此类技术不仅消耗了攻击者的时间和资源,延缓了攻击进程,还为防御方争取了宝贵时间,以全面分析攻击、修复漏洞、强化防御。

3 基于人工智能技术的网络空间安全防御策略建议

3.1 建设智能监控与预警平台

在网络安全防御体系中,智能监控与预警平台是基石,它深度融合人工智能技术,负责感知、判断和响应各类安全威胁。该平台通过无缝对接网络设备和系统,实时捕获多维度数据,如网络流量、系统日志和用户行为等。利用大数据分析和机器学习,平台整合、清洗并分析数据,形成行为画像库,实现精准预警和风险评估。数据模型需具代表性和持续学习能力,以适应网络环境变化,提升预测精度。智能监控与预警平台的核心在于自动化与智能化运维,具备自我学习和进化特性,实时调整安全策略^[18]。平台动态设定告警阈值,避免过度报警或漏报,并自动触发响应动作,与 SIEM、SOAR 系统联动执行防御策略。平台兼容性强,统一管理各类设备、操作系统和云服务,实现全网监控与协同防御。通过单一管理界面,管理员可以全局掌控网络状态,直观查看关键指标,实现全面覆盖和深度透视。预警功能要求实时、准确、操作性强,提供详细威胁描述、处置建议和操作指南,指导管理员迅速应对。平台生成威胁分析报告,为调查取证和策略优化提供依据。功能强化时,需遵循法律法规,保护用户隐私。

3.2 采用 AI 驱动威胁检测与响应机制

面对日益严峻的网络安全形势,构建以 AI 为核心的威胁检测与响应机制至关重要。该机制利用深度学习和机器学习,对网络风险进行精准、实时、大规模的挖掘与对抗。AI 驱动的威胁检测系统深入解析网络和数据,利用深度神经网络模型识别复杂攻击手法,并结合机器学习技术适应新型威胁,自动更新检测模型^[19]。通过对历史攻击样本的学习,精准捕捉新型变种和隐匿威胁,实时监测网络行为,预判攻击手法与路径。检测到威胁后,AI 驱动的自动化响应机制迅速启动,执行隔离、溯源、协调防御等措施。系统确认威胁即隔

离感染源、切断传播途径,追溯攻击路径并定位源头。AI 协调各类安全资源,根据威胁动态调整安全策略,与 SIEM、SOAR 系统深度集成,实现事件关联、优先级排序、工作流自动化,构建智能运营中心。系统自动生成事件报告,通过自然语言处理和知识图谱技术辅助决策者快速理解威胁态势。AI 驱动的威胁检测与响应机制具有持续学习与自我优化能力,能接收新情报、漏洞信息和用户反馈,通过在线学习与模型迭代优化检测算法与响应策略,支持跨组织、跨行业的威胁情报共享与协同防御。通过 AI 驱动的威胁情报平台可实现全球视野下的联防联控。

3.3 加强身份认证与访问控制的智能化

智能身份认证系统结合 AI 技术,通过生物特征识别和行为分析提供多维度、动态化的验证手段。利用深度学习模型匹配指纹、面部、声纹等生物信息,结合用户行为特征如键盘输入和鼠标轨迹,实现安全、便捷的无密码认证^[20]。基于用户画像与行为分析的智能访问控制,确保权限配置与用户角色、工作情境、行为习惯相匹配,通过 AI 动态调整权限,避免过度授权和权限滞后。智能身份认证与访问控制系统具备实时监控和智能响应能力,协同其他安全组件构建全方位防护体系。AI 算法实时监控用户登录和访问行为,识别风险事件时立即触发二次验证和权限限制,与防火墙、入侵检测、数据防泄漏等系统融合,形成统一安全视图,精准响应安全事件。系统通过在线学习更新 AI 模型,优化认证和访问控制策略,分析误报率、漏报率和用户满意度,调整认证阈值和响应规则,寻求安全与用户体验的平衡,推动策略迭代和防御生态的改善。

3.4 实施持续的智能漏洞管理和预防

智能漏洞管理利用 AI 技术全面扫描系统,精准定位潜在安全隐患。AI 驱动的漏洞扫描工具结合机器学习算法,实时监控设备状态,剖析系统,捕捉异常信号,及时发现漏洞^[21]。通过对接全球安全情报网络,获取最新威胁情报。AI 显著缩短漏洞发现时间,降低攻击成功率,保护信息系统安全。在识别漏洞后,智能漏洞管理系统使用 AI 进行漏洞评分和优先级排序。AI 算法综合评估漏洞类型、复杂度、影响范围等因素,快速筛选高风险漏洞,指导安全团队优先修复,确保突发风险得到及时遏制。平台自动追踪修复进度,动态

调整优先级,以确保响应工作有序地进行。智能漏洞管理注重持续性和预防性,构建主动防御体系。AI 实时监测环境变化,主动识别新漏洞,触发自动防护措施,如自动打补丁、调整配置、强化访问控制,预防恶意利用等。同时,AI 会预测未来漏洞发展趋势,揭示攻击模式,为前瞻性安全策略提供依据,降低攻击风险。AI 还会分析员工行为数据,识别不安全操作,提供个性化安全培训,推广最佳安全实践,如定期更新密码、启用双因素认证,减少人为疏忽引发的安全事件,强化网络安全的关键环节。

3.5 建立智能蜜罐与欺骗防御系统

智能蜜罐与欺骗防御系统利用虚假资源吸引并捕获黑客,揭示其攻击手法和动机,为防御者提供情报并增强其防护能力。AI 技术使智能蜜罐实现高度仿真和精细模拟,具备智能监控、分析与响应能力,成为多层次防御体系的关键一环。通过深度学习和机器学习,智能蜜罐能模拟复杂的 IT 场景和细微行为特征,迷惑攻击者,使其暴露攻击手段和意图。AI 赋予智能蜜罐强大的实时监控与分析能力,能够即时捕捉并解析攻击行为,识别攻击者的工具、技术和战术,揭示其动机和技术水平,提供防御策略参考。将智能蜜罐与现有安全措施结合,创建虚实交织的网络架构,引导攻击者攻击虚假目标,延缓其对关键资源的攻击,争取响应时间。AI 技术还能自动触发诱捕与反制策略,对攻击者行为进行全程追踪,与防火墙、入侵检测系统和威胁情报平台深度集成后,可实现协同防御。使用智能蜜罐需结合具体业务场景和安全需求定制部署,确保高度仿真和安全隔离,避免被利用为跳板攻击内部网络;另还需要定期评估其有效性,通过攻击模拟和红蓝对抗检验其能力,持续优化配置和响应策略。

4 应用案例

某金融机构正在遭受频繁的网络攻击,包括高级持续性威胁(APT)、勒索软件和 DDoS 攻击。为了应对这些威胁,该机构决定采用一系列基于 AI 的安全防御措施,包括机器学习驱动的入侵检测系统、深度学习的威胁情报分析、自动化响应与防御策略优化、NLP 在安全事件关联分析中的应用、预测性维护与漏洞管理以及智能蜜罐与欺骗防御技术。AI 安全技术应用效果如表 1 所示。

表 1 AI 安全技术应用效果对比

技术名称	应用效果	攻击检测率/%	假警报率/%	响应时间	预防能力
ML-based IDS	自动检测 0 Day 攻击,有效识别未知威胁	90	5	快速	中
深度学习威胁情报分析	从海量数据中提取隐蔽攻击模式,提高智能防火墙效率	85	8	较快	强
自动化响应与防御策略优化	快速、精准地应对各类威胁,减少人工干预	95	2	实时	强
NLP 安全事件关联分析	提升事件识别准确率,增强防护系统综合性能	92	4	较快	中强
预测性维护与智能漏洞管理	降低老化漏洞引发的风险,及时修复系统漏洞	88	3	中	强
智能蜜罐与欺骗防御技术	耗费攻击者资源,延缓攻击进程,为主动防御创造机会	93	1	持续	强

通过表 1 可以看出,不同的 AI 安全技术在攻击检测率、假警报率、响应时间和预防能力方面各有优势。例如,自动化响应与防御策略优化在响应时间上表现最为突出,而智能蜜罐与欺骗防御技术在预防能力方面表现出色。该金融机构可以根据自身需求和威胁类型,选择最适合的技术组合,构建多层次、智能的防御体系,有效抵御网络攻击。

5 结语

在网络安全领域当中,面对不断涌现的新威胁,需要高度关注以及应对全面运用人工智能所形成的双刃剑效应,不仅需要借助此类技术对安全防线进行完善,同时还必须高度重视内在风险。因此,必须对以 AI 技术为基础的各类防御技术进行持续研发和完善,构建一整套科学高效的智能预警机制,不断加强用户身份认证以及访问控制,同时借助诸多新技术手段来开展防御。

参考文献:

[1] 段晓东,李婕妤,程伟强,等. 面向智算中心的新型以太网需求与关键技术[J]. 电信科学,2024,40(6):146-159.

[2] 苏宇. 公共数据污染的法律治理[J]. 交大法学,2025,16(2):50-63.

[3] 王聪丽,李金慧,王靖然,等. 5G 网络量子安全算法应用研究[J]. 电信科学,2024,40(11):135-147.

[4] KUMA S , KUMAR D , DANGI R , et al. A Review of lightweight security and privacy for resource-constrained IoT devices [J]. Computers, Materials & Continua, 2024, 78(1): 31-63.

[5] 扈红超,隋嘉祺,张帅,等. 网络安全主动防御技术:策略、方法和挑战[J]. 计算机科学,2024,51(增刊2):829-841.

[6] 何清林,王丽宏,陈艳姣,等. 一种自动实时的物联网在野漏洞攻击检测方法[J]. 北京航空航天大学学报,2024,50(7):2195-2205.

[7] KOSTAS K, JUST M, LONES M A. Individual packet

features are a risk to model generalisation in ML-based intrusion detection [J].IEEE Networking Letters, 2025, 7(1):66-70.

[8] 沈华,田晨,郭森森,等. 基于对抗性机器学习的网络入侵检测方法研究[J]. 信息安全学报,2023,23(8):66-75.

[9] 冉晴月,林伟,杨知方,等. 基于可信深度神经网络的最优潮流计算方法[J]. 电工技术学报,2024,39(21):6687-6699.

[10] 赵洪强. 网络入侵检测与入侵响应技术研究[J]. 中国新通信,2025,27(3):45-47.

[11] 王耀祖,李擎,戴张杰,等. 大语言模型研究现状与趋势[J]. 工程科学学报,2024,46(8):1411-1425.

[12] 薛亮,汪含笑,胡晓波,等. 基于用户实体行为分析与人工智能的数据安全审计管理系统研究[J]. 信息通信技术与政策,2024(12):73-81.

[13] 金立杰,许旭,吴盛威,等. 深度学习驱动的设备预测性维护现状与展望[J]. 制造业自动化,2025,47(2):123-131.

[14] 王瑞,阳长江,邓向东,等. 欺骗防御技术发展及其大语言模型应用探索[J]. 计算机研究与发展,2024,61(5):1230-1249.

[15] 杨宏宇,谢丽霞. 网络入侵诱骗技术:蜜罐系统的应用[J]. 计算机工程,2006,32(13):176-178,181.

[16] 韩凤董,宗学军,何戡,等. 面向网络安全不平衡数据的特征学习和分类研究应用[J]. 科学技术与工程,2023,23(3):1130-1137.

[17] 刘亚群,高雅卓,邢长友,等. 面向欺骗防御的蜜网技术研究[J]. 计算机应用与软件,2024,41(10):362-371.

[18] 杨艳琴. 深度学习在金融网络安全监测与预警系统中的应用研究[J]. 网络安全和信息化,2025(2):55-57.

[19] 李翠红. 基于 AI 的网络安全态势感知模型与实时响应策略[J]. 计算机应用文摘,2024,40(24):173-175.

[20] 朱金涛,魏银珍,尚晓晓. 基于区块链的去中心化动态身份认证系统[J]. 计算机应用与软件,2025,42(1):333-337.

[21] 刘尚麟,陈怡多,黎珂. 协作自主 AI Agent 系统及工业系统网络安全应用研究[J]. 工业信息安全,2023(3):61-67.

收稿日期:2024-07-01