 DOI: https://doi.org/10.12345/bdai.v6i3.28259

Research on Network Security Defense Mechanism Based on Artificial Intelligence

Fang He[1] Lei Yu[1] Yanqiang Li[2] Jingxin Wu[2]

1. China Mobile Communications Group Guangdong Co., Ltd. Guangzhou Branch, Guangzhou, Guangdong, 510220, China
2. China Mobile Communications Group Guangdong Co., Ltd. Foshan Branch, Foshan, Guangdong, 528000, China

Abstract

The rapid development of artificial intelligence provides technical support for the construction of network security defense mechanisms. Based on this, the article first analyzes the deficiencies of the traditional network security epidemic prevention mechanism, and then explores the construction strategies of the network security defense mechanism based on artificial intelligence around four dimensions: information security detection, data storage and disaster recovery backup, system access permission control, and network security early warning and emergency response, and gives optimization suggestions.

Keywords

Artificial intelligence; Cyber security; Defense mechanism

基于人工智能的网络安全防御机制研究

何方[1] 郁磊[1] 李炎强[2] 吴靖欣[2]

1. 中国移动通信集团广东有限公司广州分公司，中国·广东 广州 510220
2. 中国移动通信集团广东有限公司佛山分公司，中国·广东 佛山 528000

摘　要

人工智能的迅猛发展，为网络安全防御机制建设提供了技术支持。文章基于此，首先分析了传统网络安全防疫机制的不足之处，继而围绕信息安全检测、数据存储与容灾备份、系统访问权限控制、网络安全预警与应急四个维度，探讨了基于人工智能的网络安全防御机制构建策略，并给出了优化建议。

关键词

人工智能；网络安全；防御机制

1 引言

网络安全是国家安全的重要组成部分，加强网络安全防御机制建设，既是防范网络安全风险，营造清朗网络环境的内在要求，也是保障国家安全的客观需要。人工智能以模拟和延展人类智能为主要特征，能“把人的一部分智能活动通过机械化的形式表现出来”[1]。对网络安全而言，人工智能是一柄“双刃剑”。一方面，人工智能使得网络攻击手段呈现出智能化、隐蔽化、规模化的特点，如利用人工智能生成钓鱼邮件、通过生成对抗网络（GAN）制造绕过传统检测的对抗样本等，极大地增加了网络安全的防御难度，另一方面，人工智能也为新型网络安全防御机制建设提供了技术支持。因此，要构建基于人工智能的网络安全防御机制，全面提升网络安全风险防范能力。

【作者简介】何方（1983–），男，中国吉林舒兰人，硕士，高级工程师，从事IT开发和网络安全研究。

2 传统网络安全防御机制的局限性分析

人工智能背景下，网络安全风险日益复杂化，各类型的网络安全事件时有发生。传统的网络安全防御机制，在网络安全风险，特别是新型网络安全风险的应对中，存在着很大的局限习惯。首先，规则依赖与滞后性。传统入侵检测系统（IDS）高度依赖 Snort 规则集等预定义的攻击特征库，需人工更新规则以应对新攻击。但零日漏洞（Zero-day）攻击从发现到特征提取的平均周期为 21 天，此期间防御系统完全失明，难以起到网络安全防御的作用，加剧了网络安全风险。其次，高误报率与资源消耗。传统的网络安全防御系统，通过阈值匹配检测异常，短时间内 TCP 连接数超阈值，批量下载等合法操作易触发误报，影响安全预警的准确性，并使得大量资源并用于无效警报的验证，造成资源消耗。最后，复杂场景适配性差。传统网络安全防御机制难以适配智能摄像头、工业传感器等物联网设备产生的低带宽、短包等非结构化流量以及云环境中的多租户混合流量，致使安全防

御难以覆盖所有场景，安全管理存在漏洞。人工的核心优势在于从海量数据中自主学习模式，这与网络安全的数据密集型需求高度契合，而机器学习、深度学习、强化学习等技术可高效完成特征提取、异常检测与决策优化，为基于人工智能的网络安全防御机制建设提供了良好条件。

3 基于人工智能的网络安全防御机制构建策略

3.1 以人工智能完善信息安全检测机制

网络攻击会导致系统信息泄露、丢失、篡改等安全事件，严重威胁网络安全。应用人工智能完善信息安全检测机制。针对不同形态的文档，如文本类、图像类、音视频类等，采取差异的检测方式，做好精准识别。以文本类档案为例，内容修改、文本替换、数据更替等，是文本类最常见的篡改方式。可利用 BERT 预训练模型提取文本语义特征，对比不同版本文档的语义向量差异，识别语义矛盾，结合 OCR 技术对扫描件文字进行二次识别，从而精准识别出篡改痕迹。针对个人身份证号、社保账号、经营数据等重要信息的泄露风险，可利用人工智能技术创新脱密方式，防范信息泄露风险。先基于语料库，微调 BERT 模型，识别身份证号等实体，再结合语义理解判断敏感信息的关联性，根据文档使用场景选择掩码、哈希、泛化等脱敏方式。

3.2 以人工智能创新数据存储与容灾备份机制

数字时代，由“0，1”构成的数字代码，成为信息的主要存储形式，数据安全也成为网络安全的重中之重。数据蕴藏着巨大的价值，介质失效或灾难导致的数据丢失，严重威胁数据安全。应围绕存储介质健康监测与智能容灾备份两个维度，加强人工智能技术的应用。从储介质健康监测的角度而言，硬盘、磁带等介质，有一定的使用年限。以往，存储介质的健康监测，主要依赖固定周期检测，无法及时发现早期故障。应利用人工智能技术，遵循数据采集、模型训练、动态调度的流程，构建预测模型，实现存储介质健康状态的实时评估。先采集存储介质读写延迟、坏扇区数量等 SMART 日志以及环境、温度等参数，利用 LSTM 模型，预测介质剩余寿命，当预测某硬盘故障率超过阈值时，自动触发数据迁移至备用存储。从智能容灾备份的角度而言，传统的容灾备份，采用的主要是全量备份+增量备份的固定策略，难以根据数据重要性动态调整。可利用人工智能技术构建基于重要性的智能容灾备份机制。将访问频率、关联业务影响度、保密等级等作为数据重要性评估的指标，以量化评分的形式，得出不同数据的重要性得分，确定备份的优先级，再利用地震、网络攻击等历史灾难数据，模拟风险场景，训练模型，预测不同区域 / 存储节点的风险概率，对重要性高、安全风险大的重点数据，采用实时同步 + 多副本异地存储的模式，而对重要性一般、安全风险可控的数据，采用定期增量备份的模式。

3.3 以人工智能技术优化系统访问权限控制

非法访问、越权访问，是网络系统常见的安全风险。基于身份认证的访问权限控制，则是遏制非法访问、越权访问的主要手段。以往，身份认证多采用用户名及密码认证的认证形式，用户在网络系统登录界面输入用户名、密码，若用户名、密码与数据库收录的用户名、密码匹配，则认证通过。用户名及密码认证较为简单，被破解的风险较高。人工智能，特别是语音识别技术、表情识别技术、行为识别技术的发展，推动了身份认证方式的创新。MFA（Multi-factor Authentication）、FI（Federated Identity）、SSO（Single Sign-On）等新型身份认证方式，具有更高的安全性。以 MFA 为例，其为基于多重因素的身份认证方式，用户在系统登录时，不仅要输入用户名、密码，还有提供其他能够证明身份的信息，如面部、指纹、硬件令牌、短信验证码。任何一项信息错误，均无法使用平台 [2]。可采用基于人工智能的新型身份认证方式，提高非法访问、越权访问风险的防御能力。以往，基于身份认证的访问控制，仅根据用户角色分配权限，无法适应临时任务、跨部门协作等场景。可利用人工智能技术，结合用户属性、环境属性、数据属性，动态调整访问权限，如检测到用户在非办公地点登录时，自动限制其访问重要数据。

3.4 以人工智能构建网络安全预警与应急机制

安全风险的发生，多伴有非授权 IP 登录、短时间内大量下载等异常行为，这为安全风险的预警提供了支持。可利用人工智能技术，围绕用户行为基线、异常检测、威胁分类构建网络安全预警机制。先采用 Autoencoder（编码器）学习用户访问时间、文件类型、操作频率等正常行为模式，形成用户行为基线，再于网络系统中引入异常检测模型，当实时行为与基线的偏离度超过阈值，触发预警，同时，结合威胁知识库，对异常行为分类，如非法访问、越权访问、恶意攻击等，为安全措施的采取提供依据。网络系统作为数字时代的新型基础设施，在组织管理、业务开展、决策咨询等方面，发挥着重要的作用。网络安全事件一旦发生，将造成难以估量的损失。可通过人工智能技术，构建基于知识图谱与强化学习的应急机制。先整合网络系统拓扑、历史事件、处置方案等数据，形成安全风险、事件场景、应急措施的关联网络，再模拟数据泄露、非法访问、恶意攻击等安全事件，开展强化学习训练，由训练模型选择适配度最高的处置路径，生成处置建议，人工确认后，自动执行。基于人工智能技术的预警与应急机制，既能将网络安全防御的重点从事后应对转变为事前防范，也能最大限度降低安全事件的负面影响。

4 基于人工智能的网络安全防御机制优化路径

4.1 优化模型性能

基于人工智能的网络安全防御机制，核心是在网络系

统中构建人工智能模型，发挥模型在风险防控与智能管理中的作用。模型的脆弱性，会影响安全防御机制的效果。举例而言，OCR 识别、内容分类模型等易受对抗样本干扰，将保密信息误判为公开信息，威胁信息安全。模型训练阶段，应注入对抗样本，提升模型对异常输入的判别能力，比如，OCR 模型的训练中，利用 FGSM（快速梯度符号法）生成对抗样本，训练 OCR 模型识别干扰后的网络文字的能力。引入 LIME、SHAP 等可解释人工智能方法，为模型决策提供理由。人工智能模型标记某份网络为敏感时，可输出敏感词、置信度，辅助人工审核。同时，采用 Federated Learning（联邦学习）技术，在不传输原始网络数据的前提下，通过模型参数交换实现跨机构联合训练，如不同单位在本地训练模型，共享加密后的梯度信息 [3]。

4.2 注重技术的多维协同

除人工智能技术外，其他数字技术，在网络安全风险防御中也有着重要的应用价值。以区块链技术为例，区块链（Blockchain）是记录信息生成的区块组成的链条，最早由日本人中本聪（Satoshi Nakamoto）提出，具有去中心化、分布式存储、智能合约、信息不可篡改等特点，是网络安全的可靠保障。技术的协同应用，能够进一步发挥人工智能在网络安全防御中的作用。首先，人工智能技术与大数据技术的协同。大数据技术具有极强的数据分析能力，不仅可以分析结构化数据、半结构化数据以及非结构化数据，且能通过关联分析、聚类分析、偏差分析等工具，挖掘数据价值。应围绕数据结构、数据数值、数据内容、数据交换四大方面，出台统一的网络数字化标准，提升网络数据的规范性，为人工智能算法的应用创设良好的条件。其次，人工智能技术与区块链技术的协同。区块链技术分布式存储以及信息不可篡改的特点，使得其在网络的数据存储以及防篡改中有着良好的应用效果。可利用区块链技术的分布式账本及哈希算法，将数据以区块的形式存储于区块链平台中。每个区块均包含一区块的哈希值，形成链式结构。数据上链后，任何信息篡改，均会导致哈希值变化。理论上，需超过 51% 节点同意，方可篡改。联盟链场景下，这一条件不具实现的可能性，从而极大地保证了数据的原始性和真实性。

4.3 构建新型安全治理体系

人工智能技术的应用，从根本上重塑了网络安全管理的模式，需要构建新的治理体系。当前，治理体系中面临着不少的问题。过度依赖 AI 自动审核可能导致人工复核机制失效。举例而言，如果人工智能模型长期未更新，对新型敏感内容识别失效，可能引发技术性盲区。又如，责任界定困难。导致网络安全事件的因素有很多，如模型缺陷、数据问题、人为操作，责任追溯机制的缺失，不利于判定网络安全事件的责任归属。应以制度创新为抓点，构建新型安全治理体系。首先，设立人机协同的双轨制流程。人工智能技术能够部分取代安全员的工作，但不能完全替代安全员。许多场景下，人工智能技术仅起辅助作用，最终权限仍由网络员掌控。可构建 AI 初筛 + 人工终审的双轨制流程，比如，AI 标记为敏感的网络需经 2 名以上安全员人工复核，标记为非敏感的网络按一定比例随机抽检，避免技术依赖导致的盲区。其次，构建责任链追溯机制。针对人工智能技术应用引发的新风险，可构建覆盖模型开发者、数据提供者、系统运维者三方的责任链追溯机制，针对安全事件的表现、成因，开展责任溯源，如因模型未更新导致敏感内容漏判，需追溯模型版本管理记录，界定开发者是否未及时优化。

5 结语

当前，数据泄露、篡改伪造、存储介质失效、非法访问等网络安全风险频发。传统依赖人工审核、规则匹配的安全防护手段已难以应对动态化、隐蔽化的威胁。推动网络安全防御机制创新，成为加强网络安全管理的关键。对此，应深刻认识到人工智能在网络安全防御中的作用，构建基于人工智能的网络安全防御机制。

参考文献

[1] [英]玛格丽特·A·博登.人工智能哲学[M].刘西瑞，王汉琦译.上海：上海译文出版社，2006：72.

[2] 陈曦.基于人工智能技术的计算机网络安全防御系统设计[J].数字通信世界，2025（01）：100-102.

[3] 尹智.基于大数据及人工智能技术的计算机网络安全防御系统构建研究[J]. 华东科技，2024（06）：92-94.