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

攻防对抗视角下网络安全主动防御体系研究

林飞
山东大学
DOI:10.12238/acair.v3i2.13526

[摘　要] 面向日益严峻的网络攻防对抗态势,在可信数据空间中保护人工智能系统安全成为重要挑战。传统纵深防御因防御能力固化、缺乏灵活性,难以应对高级持续性威胁等动态攻击。为此,本文研究一种面向大数据基于强化学习的自主演化大模型驱动主动防御体系。首先,构建主动防御的理论建模框架,将攻防过程建模为动态博弈,并采用深度强化学习求解最优防御策略,实现防御策略的持续优化与进化。其次,设计主动防御系统架构,融合大数据分析平台,利用大模型实时感知威胁、决策防御响应,并通过标准接口集成国产软硬件基础设施。在AI系统攻击场景进行对抗性实验,训练防御智能体自动识别并挫败攻击。
[关键词] 网络安全；攻防对抗；主动防御
中图分类号：TV　文献标识码：A

Research on Proactive Cybersecurity Defense Systems from an Attack–Defense Perspective

Fei Lin
Shandong University

[Abstract] In the face of an increasingly severe cyber attack and defense landscape, securing artificial intelligence systems within trusted data spaces has become a significant challenge. Traditional defense-in-depth strategies, due to their rigid defense capabilities and lack of flexibility, struggle to cope with dynamic attacks such as Advanced Persistent Threats (APTs). Therefore, this paper investigates an autonomous evolutionary large model-driven active defense architecture based on reinforcement learning for big data. Firstly, a theoretical modeling framework for active defense is constructed, modeling the attack-defense process as a dynamic game and employing deep reinforcement learning to solve for the optimal defense strategy, thereby achieving continuous optimization and evolution of defense strategies. Secondly, an active defense system architecture is designed, integrating a big data analytics platform and leveraging large models for real-time threat perception and defense response decision-making, while also integrating domestic hardware and software infrastructure through standard interfaces. Adversarial experiments are conducted in AI system attack scenarios to train the defense agent to automatically identify and thwart attacks.
[Key words] Network security; Attack and defense confrontation; Active defense

引言

随着人工智能(AI)技术融入关键业务和基础设施,针对AI系统的网络攻击不断演进,呈现对抗性和持续性的特点。传统网络安全主要依赖纵深防御策略,通过防火墙、入侵检测系统等多道静态防线来抵御攻击[1]。然而,高级持续性威胁(APT)常利用零日漏洞和隐蔽手段潜伏于系统,传统固定防御难以及时发现和响应。例如,SolarWinds供应链攻击和对深度学习模型的对抗样本攻击显示出攻击者可以绕过静态防御并长期潜伏,因此亟需更主动智能的防御体系[2]。

主动防御理念由此兴起,强调防守方采取动态、多变的策略与攻击者周旋[3]。主动防御不同于被动防御在于：不仅被动拦截已知攻击,还主动诱骗、迷惑和阻断未知攻击。近年来,主动防御具体策略包括移动目标防御(Moving Target Defense, MTD)、网络诱骗(如蜜罐技术)以及自主响应等[4]。研究表明,持续变换IP地址、端口等MTD机制可显著削弱攻击效果,模拟结果表明攻击成功率随防御变化频率提高而降低[4]。

与此同时,网络防御逐渐引入博弈论和机器学习等方法提升智能化水平。攻防对抗本质上可抽象为动态不完美信息博弈,防守方需在不确定环境下与对手反复博弈。早期研究使用静态博弈分析最优防御策略选择。例如,Zhang等构建了静态贝叶斯博弈模型求解最优防御策略。随后,动态博弈模型被用于描述多步攻击过程,如微分博弈和马尔可夫博弈模型,用以分析攻击-

防御在时序上的策略演化[5]。Huang等将网络攻防建模为微分对策,求解状态转移随机的Markov微分博弈均衡策略[6]。这些博弈模型为主动防御策略优化提供了理论基础,但基于博弈求解需要已知准确的对抗模型,在真实复杂环境下存在局限。近年来,强化学习(Reinforcement Learning,RL)凭借在不完备信息下通过试错学习最优策略的能力,开始应用于网络对抗决策。

综上所述,本文面向AI系统安全防护,提出将大模型驱动的深度强化学习应用于网络安全主动防御体系设计,旨在解决传统防御刚性不足的问题。本文主要贡献包括：(1)在理论上,构建了主动防御攻防对抗的形式化模型,引入深度强化学习方法求解演化博弈中的最优防御策略,实现防御策略随攻击动态自适应优化；(2)在架构上,基于国产信创环境的大模型主动防御系统,研究实现大数据环境感知、策略决策与执行反馈的闭环,将国产软硬件融入防御框架；(3)在实验上,通过真实/模拟的AI系统攻击场景,对比评估所提体系在攻击检测率、响应速度等方面的效能提升,并验证强化学习防御智能体与攻击智能体对抗训练的有效性。

1 相关工作

1.1主动防御与移动目标防御。主动防御理念最早由军事防御引入网络安全领域,Jajodia等编辑的专著系统阐述了移动目标防御(MTD)的原理,将其作为改变网络攻防“不对称”格局的重要技术。MTD通过动态改变系统配置来“移动”攻击目标,使攻击者难以侦察和利用漏洞[7]。Okhravi等对已有MTD技术进行了综述,包括随机化地址、指令集多样化、虚拟化迁移等手段[8]。Lei等指出MTD作为一种主动防御技术,近年成为研究热点,能有效提升攻击方的不确定性和攻击成本[9]。在实践中,Sharma等将MTD与软件定义网络结合,实现了IP地址重用和路由随机化,有效防御针对物理基础设施的侦察攻击[10]。

1.2攻防博弈建模。将攻防互动形式化为博弈模型是主动防御理论研究的重要方向。早期工作如Lye和Wing采用静态博弈分析入侵者与管理员的策略均衡[11]。随后,动态博弈模型兴起,包括Stackelberg博弈用于模拟进攻与防守的先后次序,FlipIt博弈用于描述隐蔽攻击场景下资源控制的轮换[12]。Van Dijk等提出的FlipIt游戏模型刻画了攻击者反复秘密夺取系统控制权、防守方偶尔检查并重置的动态过程,为APT检测策略提供了新思路[13]。在FlipIt框架下,Chen等结合Q-learning优化APT检测的时机,提出基于FlipIt的APT主动探测策略,使防守方通过学习对手入侵间隔来选择最优扫描频率。

1.3强化学习在网络攻防中的应用。深度强化学习的兴起为主动防御策略优化带来了新机遇。与传统方法不同,RL智能体通过与环境交互自主学习,而非依赖预设规则,对于动态多变的网络攻击非常契合。Nguyen和Reddi详细综述了RL在网络安全中的研究进展,表明RL已被用于入侵检测、恶意软件分类、入侵响应、移动目标防御等多个领域[14]。在入侵响应方面,Hammar和Stadler将入侵防御视为最优停止问题,通过深度强化学习智能体学会何时阻断连接以最小化业务中断和安全风险。其方法使防御系统在攻击发生时并非立即切断服务,而是根据攻击严重度和对业务影响的权衡来决定响应时机,从而平衡了安全与可用性。

1.4大模型与网络安全。随着Transformer等预训练模型的出现,大模型也开始用于网络安全。一方面,大模型可以作为高级威胁情报分析和安全运维的助手。Hassanin和Moustafa概述了大型语言模型(LLM)在威胁检测、事件响应、安全自动化等方面的应用潜力,认为LLM凭借对大数据的理解能力,可用于识别复杂攻击模式和生成防御建议。另一方面,大模型自身也可能成为攻击目标或攻击手段。Alblehai等在物联网安全研究中指出,攻击者正利用生成对抗网络(GAN)和LLM来自动化发动网络攻击(如生成钓鱼文本、自动化漏洞利用),这使得防守方面临AI驱动的新型威胁。因此,将大模型用于防御既有巨大机遇也有挑战。一些工作尝试将安全领域知识融入预训练模型,从而增强其检测未知威胁的能力。

2 理论方法

2.1攻防对抗模型建模。将网络攻防过程抽象为一个Markov决策过程(MDP)$< S, A_D, A_A, P, R_D, R_A >$,其中状态空间S刻画了系统安全状态及攻击者信息,可由安全事件日志、大数据分析结果综合表示。例如S可包括当前系统配置、安全警报级别、攻击者可疑行为特征等。A_D和A_A分别为防御方与攻击方的动作集合。防御动作$a_D \in A_D$可能包括：调整防火墙规则、启用MTD(如更换端口/IP)、部署蜜罐诱捕、切断会话、启动应急策略等；攻击动作$a_A \in A_A$则包括侦察(扫描端口/服务)、利用漏洞渗透、横向移动、数据泄露等。双方动作的执行会引起系统状态转移,记$P(s'|s,a_D,a_A)$为状态转移概率分布。由于攻防是非合作对抗关系,防御者和攻击者具有不同的奖赏函数R_D和R_A。将防御者奖赏$R_D(s,a_D,a_A)$为状态s下采取防御动作a_D对防御方的即时收益,可定义为负的损失函数,例如攻击是否被阻止、系统性能损耗等的加权和。攻击者奖赏$R_A(s,a_D,a_A)$则与之相反,表示攻击成功与否及其收益(如窃取数据价值)等。

2.2强化学习求解优化策略。由于攻防状态转换和攻击者策略可能非常复杂难以用解析方法求解,本研究采用深度强化学习算法来近似求解最优防御策略。即,防御智能体以MDP形式与环境交互。在状态s_t下观察到安全上下文后选择一个防御动作,$a_t^D = \pi(s_t)$环境(包含潜在攻击者)随后转移到新的状态s_{t+1},并给予防御智能体一个奖赏r_t^D(同时攻击者得到

$r_t^A = -r_t^D$）。通过不断尝试不同策略并积累奖赏,防御智能体可以更新其策略以最大化长期回报。本文采用的强化学习算法包括深度Q网络(DQN)和策略梯度方法相结合,即Actor-Critic架构。

Actor网络参数化防御策略$\pi(a|s;\theta)$,输出在状态下各防御动作的概率分布；Critic网络近似状态-动作值函数$Q(s,a)$或状态值函数$V(s)$,用于指导Actor优化。智能体训练目标是极大化折扣累积奖赏$R = \sum_t Y^t r_t^D$（γ为折扣因子)。为提高训练稳定性,引入经验回放和目标网络技术,并使用近端策略优化(PPO)算法更新策略参数。

2.3大模型融合策略。为充分利用大数据情报并提高决策质量,本研究在强化学习框架中融入预训练的大规模AI模型。大模型特征提取：利用预训练模型从原始安全数据中提取高维特征作为RL智能体的输入。比如,可采用训练好的Transformer模型分析流量序列或日志文本,将复杂模式编码为向量表示输入Actor-Critic网络。这相当于构造一个功能强大的感知模块,使智能体在观察状态时得到更具语义和预测能力的特征。尤其在AI系统场景下,大模型可帮助识别对抗样本攻击的潜在意图、检测异常系统行为模式等,从而丰富状态描述。实验中我们引入了一个基于BERT的大模型对系统调用序列进行编码,强化学习智能体据此判断是否存在隐蔽攻击。大模型决策引擎：将大模型直接参与决策过程,例如采用大语言模型根据当前情境生成防御策略建议,再经过RL智能体评估采纳。

3 系统架构

3.1感知层。大数据安全态势感知。感知层负责收集和处理网络空间的海量多源数据,构建实时安全态势,为决策层提供支持。感知层由国产软硬件构成的数据采集与分析平台实现；大数据存储与计算模块,对采集数据进行聚合清洗和快速检索；威胁情报与分析模块,利用机器学习和大模型分析海量数据,提取可疑模式与告警。此外,通过引入安全数据湖,汇聚网络流量的NetFlow记录、主机的操作日志、身份认证记录等,并预置AI模型如异常流量检测模型、用户行为分析模型等。

3.2决策层。强化学习决策引擎。决策层是主动防御体系的大脑,位于可信计算环境中,运行防御智能体和策略决策逻辑。它由决策智能体模块和策略生成与评估模块构成。决策智能体模块实现了上一节所述的强化学习算法,包括Actor-Critic网络结构。其中,Actor网络由卷积神经网络和全连接层构成,用于从感知层输入的高维状态特征中提取决策所需表示并输出动作概率分布；Critic网络采用类似结构输出状态值$V(s)$估计。策略生成与评估模块则用于在训练阶段更新决策智能体策略,以及在运行阶段辅助决策。训练阶段,策略生成模块会调用对抗训练环境,与模拟攻击智能体进行大量训练对局,不断优化Actor-Critic网络参数。

3.3执行层。主动防御执行机构。执行层负责根据决策层输出的防御动作实施具体操作,直接作用于网络和主机系统。它主要包括：配置管理模块,通过编排和控制接口执行对网络设备、主机系统配置的调整,如下发防火墙规则变更、触发SDN控制器调整路由；诱骗与隔离模块,部署各种诱骗陷阱并在决策要求时激活,同时具备对受感染主机或恶意流量进行隔离封堵的能力；系统恢复模块,在攻击造成损害时执行备份恢复、补丁修复等操作,以减轻攻击影响。为此,本研究采用了基于Agent的软件执行单元部署在各节点,就近执行指令,减少集中控制带来的延迟。

4 实验设计

4.1攻防智能体训练。在对抗训练阶段,启动一个攻击智能体来模拟APT攻击者的策略演化。攻击智能体的动作空间包括扫描、利用(可选多个漏洞)、横向移动(选择下一个目标)和隐藏痕迹等,奖赏设计为成功控制关键服务器则获得高奖赏,过早暴露导致被封堵则获得负奖赏。本研究采用深度Q学习训练攻击智能体,使其逐步学会更隐蔽高效的入侵策略。在此过程中,防御智能体则通过PPO算法不断优化防御策略以应对攻击智能体。

4.2对比方案设置。为了评估体系性能,本研究设计了多组对比实验：(a)静态防御：不采用主动防御策略,仅依赖传统安全设备(防火墙、IDS)按固定规则防御；(b)规则自适应防御：采用一定简单自适应策略,如基于阈值触发的MTD,即当攻击告警数超过阈值时切换IP(这代表传统安全产品的自动化响应能力)；(c)强化学习防御：即本文提出的RL驱动主动防御体系；(d)RL防御+大模型：在(c)基础上增加LLM威胁分析辅助决策,评估大模型的增益。针对每种方案分别进行了多次攻击模拟,并记录关键指标。

4.3评估指标。从安全效能和业务开销两方面评估。(1)攻击成功率:攻击者在每轮攻击中是否达到主要目标(控制客服服务器或使AI模型输出错误关键回复)。该指标反映防御有效性,越低越好。(2)攻击检测延迟：从攻击开始到防御体系首次识别出攻击的时间。(3)响应时间：从识别攻击到采取防御动作的延迟。(4)业务可用性：在整个攻击防御过程中,正常服务是否中断或性能下降,比如客服响应时间、请求成功率变化等。(5)防御策略改变频率：主动防御采取动作的频度,如MTD切换IP的次数等,作为防御开销的衡量。

5 结果分析

5.1攻击成功率与检测效果。在基于静态防御的方案,APT攻击成功率高达60%,对抗样本攻击成功率约75%,表明传统静态防线难以阻止巧妙攻击。简单规则自适应方案有所改善,将APT成功率降至40%,但对抗样本攻击仍有近70%成功率,原因在于此方案主要对网络入侵做出响应,而对AI模型的异常输入缺乏机制。相比之下,强化学习防御方案将APT成功率降至15%,对抗样本成功率降至20%,显著优于前两者。这表明RL智能体学到了更有效的综合防御策略,例如在APT早期侦察阶段即发现端倪并采

取措施,中断了攻击链；同时对对抗样本攻击,智能体通过多特征关联分析识别出异常查询并进行过滤。当引入大模型辅助后,攻击成功率进一步略降至APT10%、对抗攻击15%。

5.2检测率与响应及时性。在攻击检测方面,RL方案表现出色。统计显示,静态方案对APT的平均检测延迟约120秒,因为只有当攻击进入明显阶段才被传统IDS捕获;RL方案将检测延迟缩短到45秒,大模型辅助进一步缩短至30秒左右。这是因为强化学习智能体能够结合多维异常特征,在攻击早期阶段就发现异常并积累证据,而大模型提供的上下文知识让智能体更快做出判断。对于对抗样本攻击,静态方案几乎检测不到,RL智能体利用模型输入输出的相关特征,实现了约1秒内识别异常输入。响应延迟方面,RL方案在检测后平均2秒内执行防御动作,远低于静态方案人工响应往往需分钟级别。

5.3策略实例分析。为了更直观地理解强化学习防御策略,我们提取了一些典型对抗过程加以分析。在一次APT攻防模拟中,攻击者首先对多台服务器进行低频扫描。传统防御未发觉异常,但我们的RL智能体捕获到多个终端同时出现少量异常端口访问这一罕见模式,推测可能是分布式扫描,遂调用LLM询问建议。LLM结合历史知识返回“可能的侦察活动,应收紧访问控制”,智能体据此下发策略将临时提高防火墙敏感度并开启详细流量监控。不久攻击者对客服服务器尝试Exploit,由于防火墙策略收紧,此次利用被IPS及时阻断并告警。智能体检测到高危告警,立即触发MTD,将客服服务器IP移至新段并启用交互蜜罐在原IP处。

5.4讨论与展望。本研究设计的主动防御体系在对抗模拟中仍存在一些局限。首先,策略可解释性问题突出。深度强化学习智能体往往是“黑箱”,安全运维人员难以理解其决策依据。当智能体采取非常规动作时(如突然隔离某节点),缺乏解释可能影响对体系的信任。未来可探索引入可解释AI方法,例如对智能体的决策进行规则抽取或利用安全领域知识对策略进行约束,提高决策透明度和可信度。一些初步工作已将知识图谱应用于强化学习,以指导智能体的学习过程。其次,对抗鲁棒性值得关注。攻击者可能针对防御智能体本身发动对抗攻击,干扰其决策过程,如对采集数据投放对抗样本使智能体误判。已有研究表明深度RL模型易受对抗扰动影响,未来需考虑将对抗训练纳入防御智能体的训练流程,以提升其鲁棒性。此外,样本高效性也是挑战。训练智能体需要大量对抗样本数据,而实际中全面覆盖各种攻击场景的数据难以获取。

[参考文献]

[1]S.Ennaji,F.D.Gaspari,D.Hitaj, A. Kbidi, and L. V. Mancini, “Adversarial Challenges in Network Intrusion Detection Systems: Research Insights and Future Prospects,” Oct.22,2024,arXiv: arXiv:2409.18736.doi:10.48550/arXiv.2409.18736.

[2]Z.Tan,S.P. Parambath, C. Anagnostopoulos, J. Singer, and A.K.Marnerides, “Advanced Persistent Threats Based on Supply Chain Vulnerabilities: Challenges, Solutions & Future Directions,” IEEE Internet of Things Journal, 2025, Accessed: May 04, 2025.[Online].Available:https://ieeexplore.ieee.org/abstract/document/10838587/.

[3]Y.Yang,N.B.Idris, C. Liu, H. Wu, and D. Yu, “A destructive active defense algorithm for deepfake face images,” PeerJ Computer Science,vol.10,p.e2356,2024.

[4]D.Reti,D.Fraunholz,K.Elzer,D.Schneider,and H.D.Schotten, “Evaluating Deception and Moving Target Defense with Network Attack Simulation,” in Proceedings of the 9th ACM Workshop on Moving Target Defense, Los Angeles CA USA: ACM, Nov.2022,pp.45–53.

[5]Dingkun Yu,Tao Li,Hengwei Zhang,Jihong Han,and Jindong Wang, “Active defense strategy selection based on static Bayesian game,” in Third International Conference on Cyberspace Technology (CCT 2015), Beijing, China: Institution of Engineering and Technology,2015,p.7.

[6]S.Huang,H.Zhang,J.Wang,and J.Huang, “Markov differential game for network defense decision-making method,” IEEE Access,vol.6,pp.39621–39634,2018.

[7]A.McGibney,T.Ranathunga, and R.Pospisil, “SmartQC: An Extensible DLT-Based Framework for Trusted Data Workflows in Smart Manufacturing,” Feb.27,2024, arXiv:arXiv:2402.17868.

[8]J.-H. Cho et al., “Toward proactive, adaptive defense: A survey on moving target defense,” IEEE Communications Surveys & Tutorials,vol.22, no.1,pp.709–745,2020.

[9]R.Sun,Y.Zhu,J.Fei,and X.Chen, “A survey on moving target defense:Intelligently affordable,optimized and self-adaptive,” Applied Sciences,vol.13,no.9,p.5367,2023.

[10]D.P.Sharma,J.-H.Cho,T.J.Moore,F. F. Nelson, H. Lim, and D.S.Kim, “Random host and service multiplexing for moving target defense in software-defined networks,” in ICC 2019 -2019 IEEE International Conference on Communications (ICC), IEEE,2019,pp.1 - 6. Accessed: May 04, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8761496/

[11]K.Lye and J.M.Wing, “Game strategies in network security,” IJIS,vol.4,no.1–2,pp.71–86,Feb.2005.

[12]D.Ramsey, “A Stackelberg Game based on the Secretary Problem: Optimal Response is History Dependent,” Sep. 06, 2024,arXiv:arXiv:2409.04153.

[13]M.Van Dijk,A. Juels, A. Oprea, and R. L. Rivest, “FlipIt: The Game of ‘Stealthy Takeover,’ ” J Cryptol, vol.26,no.4,

[14]T.T.Nguyen and V.J.Reddi, “Deep reinforcement learning for cyber security,” IEEE Transactions on Neural Networks and Learning Systems,vol.34,no.8,pp.3779–3795,2021.

作者简介：

林飞(1974--),男,汉族,山东乳山人,大学本科,副研究员,研究方向为网络安全和信息化。