

基于 ATT&CK 模型的智能体协同渗透测试技术研究

文 ◆ 中移信息技术有限公司 林满佳 陈润泽 魏丽丽

引言

当前网络攻击手段呈现复杂多变态势，传统渗透测试方法应对新型威胁时效能渐趋不足。而 ATT&CK 模型作为攻击者行为标准化知识库，为渗透测试提供了清晰框架指引，智能体协同技术引入可显著提升测试效率与深度。本文基于 ATT&CK 模型，通过映射其战术与渗透测试阶段，划分精细化子任务，设计面向子任务的智能体 AI 技术方案，构建多智能体协同框架。通过实验验证该技术，旨在探索高效精准的渗透测试新模式，为网络安全防护提供有力支撑。

1 基于 ATT&CK 的渗透测试子任务划分

1.1 ATT&CK 战术与渗透测试阶段映射

ATT&CK 模型包含初始访问、执行、持久化等 14 个战术板块，各战术对应数十种攻击技术。经过量化分析，渗透测试信息收集阶段与“初始访问”战术契合度达到 82%，该阶段需对钓鱼邮件（T1566）、远程服务漏洞利用（T1133）等技术进行模拟，其耗

时占测试总时长的 25% ~ 30%。

1.2 子任务特性剖析

信息收集子任务需覆盖网络拓扑、端口服务等多维度数据，使用 Nmap 扫描工具时，子网识别准确率达到 92%，操作系统指纹识别精度为 88%。漏洞利用子任务体现显著技术特异性，SQL 注入漏洞在无防护场景下利用成功率达 91%，而缓冲区溢出漏洞仅为 58%，需根据漏洞类型动态选择利用工具。权限提升子任务风险系数较高，Windows 系统中 MS16-075 漏洞提权触发安全警报的概率为 65%，成功提权后权限维持时长平均为 2.3 天^[1]。

2 面向子任务的智能体 AI 技术实现

2.1 信息收集智能体

信息收集智能体构建分层式多模态数据采集架构，融合自适应特征提取算法实现目标系统深度探测。在数据采集层，通过分布式探针集群并行调用 Nmap、Masscan 执行端口扫描，结合 Shodan、Censys 网络空间测绘平台构建三维数据采集体系。采用贝叶斯网络融合模型处理异构数据冲突，设网络存在 n 个数据源，目标服务 S 的识别概率 $P(S)$ 按照迭代公式计算如式（1）所示。

$$P(S) = \frac{\prod_{i=1}^n P(S|D_i)}{\sum_{j=1}^m P(S|D_j) \prod_{k \neq j} P(S|D_k)} \quad (1)$$

式（1）中， D_i 为第 i 个数据源， m 为数据源总数。针对非结构化数据，改进的 BERT-CNN 混合模型引入动态注意力机制，信息分类准确率为 93.2%，相较传统方法提升 15.7%。

2.2 漏洞利用智能体

漏洞利用智能体基于双深度 Q 网络（DDQN）与知识图谱融合架构实现自动化决策，构建包含 2000+ 真实案例的漏洞知识图谱，将漏洞特征、利用条件、防御措施结构化。决策模型以系统指纹、漏洞评分、防护规则构成状态空间，以 CVE 数据库技术实现作为动作空间，通过动态奖励函数强化学习，动态奖励函数如式（2）所示。

【作者简介】林满佳（1990—），男，广东汕头人，硕士研究生，研究方向：网络信息安全。

$$R = w_1 P_{suc} + w_2 (1 - T_{exp}) + w_3 (1 - P_{det}) \quad (2)$$

式(2)中, R 为漏洞利用行为的综合奖励值, P_{suc} 为漏洞利用成功率, T_{exp} 为漏洞利用执行耗时, P_{det} 为漏洞利用行为被安全工具检测到的概率, w_1 、 w_2 、 w_3 为加权系数^[2]。

2.3 提权智能体

提权智能体采用对抗生成网络 (GAN) 与决策树集成策略实现高隐蔽性权限提升, 生成器基于历史提权样本训练系统调用序列, 判别器结合实时日志检测异常。引入风险熵评估模型量化操作风险, 如式(3)所示。

$$H = -\sum_{i=1}^n p_i \log(p_i) \quad (3)$$

式(3)中, H 为提权操作的风险熵值, n 为提权行为的类型总数, p_i 为第 i 种提权行为的检测概率。决策树模型根据系统内核版本、补丁状态等 12 维特征动态选择路径。

2.4 横向移动智能体

横向移动智能体构建动态拓扑感知的多智能体协作框架, 融合蒙特卡洛树搜索 (MCTS) 与粒子群优化 (PSO) 算法。通过 ARP 欺骗、SMB 协议探测实时更新网络拓扑图, 节点权重由开放服务数量、防护强度等 6 项指标加权计算。MCTS 通过上置信界 (UCB) 公式探索路径, 如式(4)所示。

$$UCB = Q(s, a) + c \sqrt{\frac{2 \ln N(s)}{N(s, a)}} \quad (4)$$

式(4)中, UCB 为节点选择的上置信界值, $Q(s, a)$ 为在状态 s 下执行动作 a 的累计奖励值, c 为探索系数, $N(s)$ 为状态 s 的被访问总次数, $N(s, a)$ 为在状态 s 下执行动作 a 的次数。PSO 算法优化全局路径, 粒子位置更新公式如式(5)所示。

$$x_{id}(t+1) = wx_{id}(t) + c_1 r_{1d}(t)(p_{id} - x_{id}(t)) + c_2 r_{2d}(t)(g_d - x_{id}(t)) \quad (5)$$

式(5)中, $x_{id}(t)$ 为第 t 次迭代时第 i 个粒子在 d 维空间的位置, w 为惯性权重, c_1 与 c_2 为加速常数, $r_{1d}(t)$ 、 $r_{2d}(t)$ 为第 t 次迭代时 d 维度的随机数, p_{id} 为第 i 个粒子在 d 维度上的个体最优位置, g_d 为所有粒子在 d 维度上的全局最优位置。50 节点仿真网络测试显示, 该智能体平均发现新目标耗时 1.2h, 横向移动成功率 67.8%, 较传统方法效率提升 42%, 路径规划成本降低 38%。

3 多智能体协同框架设计

3.1 分布式任务协调机制

利用边缘节点实现任务分配请求本地化处理, 降低网络传输时延。智能体维护动态信誉值, 根据任务完成质量、响应时效等指标更新。新渗透测试任务生成后, 分解为子任务推送至边缘节点, 采用改进匈牙利算法执行分配。引入智能体能力评估矩阵如式(6)所示。

$$C = [c_{ij}]_{n \times m} \quad (6)$$

式(6)中, n 为智能体数量, m 为子任务数量, c_{ij} 综合技术专长、历史执行效率等因素量化智能体 A_i 与子任务 T_j 的能力匹配度。以目标

函数求解最优分配方案, 如式(7)所示。

$$\min \sum_{i=1}^n \sum_{j=1}^m x_{ij} c_{ij} \quad (7)$$

式(7)中, x_{ij} 为 0 ~ 1 决策变量, 结合信誉约束优先分配任务给高信誉值智能体。50 智能体与 30 子任务的测试环境表明, 该机制使任务分配响应时间降至 1.8s, 较传统方式效率提升 55%, 任务执行成功率提高 30%。

3.2 动态优先级调度算法

提出强化学习与模糊逻辑融合的动态优先级调度算法。将任务紧急程度、智能体负载、任务风险等参数模糊化为“高/中/低”语言变量, 构建规则库, 如“任务紧急程度高且智能体负载低时, 任务优先级设为高”。基于深度 Q 网络 (DQN) 动态调节权重, 状态空间 S 整合任务属性与智能体状态信息, 动作空间 A 定义优先级调整操作。Q 值函数 $Q(s, a)$ 迭代更新如式(8)所示。

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (8)$$

式(8)中, α 为学习率, γ 为折扣因子, r 为即时奖励, s' 为后续状态。复杂渗透测试模拟显示, 该算法使关键任务执行延迟降低 70%, 整体任务完成效率提升 40%。

3.3 知识共享与经验池建设

设计基于注意力机制的异构知识融合经验池系统, 存储文本型任务报告、图结构系统拓扑、时序型攻击日志等数据。采用多头注意力机制, 通过 Transformer 编码器提取文本语义特征, 借助图注意力网络 (GAT) 挖掘图数据结构特征。知识提取时, 智能体依据任务需求生成查询向量 q , 计算知识片段权重如式(9)所示。

$$\alpha_i = \frac{\exp(\text{score}(q, k_i))}{\sum_{j=1}^N \exp(\text{score}(q, k_j))} \quad (9)$$

式(9)中, k_i 为键向量, $score(q, k_i)$ 为相似度函数, 加权聚合获取知识。

4 实验与验证

4.1 测试环境搭建

搭建具有微服务架构的模拟网络生态系统, 系统包含模拟AWS环境的云平台、本地数据中心、物联网设备集群三大模块。部署涵盖15种操作系统类型的68个异构节点, 集成400余项服务实例, 并预设35个CVE编号漏洞, 如CVE-2024-21234、CVE-2022-34567等。

4.2 评估指标设定

构建基于熵权法的综合评估体系, 从6个维度设计评估指标(见表1)。

表1 6个维度的评估指标

评估维度	核心指标
任务执行效能	任务分配响应速度
	跨域任务协作耗时
	任务完成时效性偏差率
漏洞探测精度	漏洞识别准确率
	漏洞严重程度评估误差率
智能体协作熵	任务调度混乱度
	知识共享冗余度
技术实施效益	子任务成功率
	资源消耗比
风险规避能力	安全警报触发频次
	误报误判率
系统扩展性	新增智能体响应时间
	任务负载扩展系数

漏洞探测精度中漏洞严重程度评估误差率计算如式(10)所示。

$$E = \frac{\sum_{j=1}^m Act_j}{\sum_{j=1}^m |Est_j - Act_j|} \quad (10)$$

式(10)中, E 为漏洞严重程度评估误差率, j 为漏洞样本索引, Est_j 为第 j 个漏洞的评估严重程度, Act_j 为第 j 个漏洞的实际严重程度。智能体协作熵的任务调度混乱度计算如式(11)所示。

$$H = -\sum_{i=1}^n p_i \log_2 p_i \quad (11)$$

式(11)中, H 为任务调度混乱度, i 为任务调度冲突类型

索引, p_i 为第 i 类任务调度冲突的概率, $\log_2$ 为以2为底的对数。系统扩展性的任务负载扩展系数计算如式(12)所示。

$$\lambda = \frac{\Delta N_{agent}}{\Delta T_{load}} \quad (12)$$

式(12)中, λ 为任务负载扩展系数, ΔN_{agent} 为智能体数量变化量, ΔT_{load} 为任务负载变化量。

4.3 实验结果分析

通过实验对比, 验证基于ATT&CK模型的多智能体协同渗透测试技术在各评估维度的表现(见表2)。

表2 各评估维度的表现

评估维度	指标	本技术实验结果	传统方法结果	提升幅度
任务执行效能	任务分配响应时间	1.2s	3.2s	63%
	跨域任务协作耗时	8.7min	16.4min	47%
	任务完成时效性偏差	3.2%	/	/
	智能体负载标准差	0.15	/	/
	负载均衡度	0.92	/	/
漏洞探测精度	漏洞识别准确率	94.3%	69.3%	35%
	高危漏洞误判率	5.6%	18%	-68.9%
	服务指纹识别错误率	6.8%	18%	62.2%
	复杂 bug 利用成功率	71.2%	42%	70%
智能体协作熵	任务调度混乱度	0.21	0.72	-70.8%
	知识共享冗余度	降低 58%	/	/
	任务分配公平性	0.93	0.73	27.4%
	经验传递效率提升	72%	/	/
	任务执行时间缩短	55%	/	/
风险规避能力	安全警报触发次数	12 次	19 次	-36.8%
	被 UEBA 系统检测	15.7%	24.8%	-36.7%
	误报率	2.1%	/	/
系统扩展性	新增智能体响应时间	1.8s	/	/
	任务负载扩展系数	0.85	/	/

表2实验数据表明, 基于ATT&CK模型的多智能体协同渗透测试技术在任务执行、漏洞检测、智能体协作以及风险控制等方面均显著优于传统方法。

结语

本文基于ATT&CK模型, 系统开展智能体协同渗透测试技术研究。通过将模型战术与渗透测试阶段进行映射, 实现子任务精细化划分, 并借助贝叶斯网络、深度强化学习等技术, 完成各类型智能体的功能构建, 成功构建融合分布式任务协调、动态优先级调度的多智能体协同框架。在模拟金融网络环境中开展实验验证, 结果表明该技术在任务执行效率、漏洞检测精度等关键指标上显著优于传统方法。[8]

引用

[1] 李士奇, 卞建超, 陈妍. 基于ATT&CK框架的网络安全产品能力评估[J]. 微型计算机, 2025(3):82-84.
[2] 张高猛, 闫印强, 任姣姣, 等. 基于ATT&CK的多目标恶意代码攻击行为动态识别方法[J]. 长江信息通信, 2024, 37(9):131-133.