

基于攻击图和深度 Q 学习网络的自动化安全分析与渗透测试模型

樊成, 胡国庆, 丁涛杰, 张展华

(中国人民解放军 31681 部队, 甘肃 天水 741000)

摘要: 随着网络技术的快速发展和广泛应用, 网络安全问题日益突出, 渗透测试成为评估和提升网络安全性的重要手段。然而, 传统的人工渗透测试方法效率较低, 且易受到人为错误和测试人员技能水平的影响, 造成测试结果不确定性大、评估效果不理想等问题。针对以上人工渗透测试中存在的问题, 提出了基于攻击图和深度 Q 学习网络 (DQN, deep Q-learning network) 的自动化安全分析与渗透测试 (ASAPT, autonomous security analysis and penetration testing) 模型。该模型由训练数据构建和模型训练两部分构成。在训练数据构建阶段, 采用攻击图对目标网络进行威胁建模, 将网络中存在的漏洞和攻击者可能的攻击路径转化为节点、边, 随后结合 CVSS (common vulnerability scoring system) 漏洞信息库构建对应的“状态-动作”转移矩阵, 用以描述攻击者在不同状态下的攻击行为和转移概率, 并全面反映攻击者的攻击能力和网络的安全状况。为进一步降低计算复杂度, 创新性地使用深度优先搜索算法对转移矩阵进行简化, 查找并保留所有能达到最终目标的攻击路径, 以便于后续模型训练。在模型训练阶段, 使用基于 DQN 的深度强化学习算法对渗透测试中的最优攻击路径进行确定, 该算法通过不断与环境交互、更新 Q 值函数, 从而逐步优化攻击路径选择。仿真结果表明, ASAPT 模型在最优路径寻找方面准确率可达 84%, 收敛速度快, 并且在面对大规模网络环境时, 相较于传统 Q 学习具有更好的适应性, 能够为实际的渗透测试提供指导。

关键词: 自动化渗透测试; 强化学习; 攻击图; 深度 Q 学习网络

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2023091

Autonomous security analysis and penetration testing model based on attack graph and deep Q-learning network

FAN Cheng, HU Guoqing, DING Taojie, ZHANG Zhanhua

The 31681 Unit of PLA, Tianshui 741000, China

Abstract: With the continuous development and widespread application of network technology, network security issues have become increasingly prominent. Penetration testing has emerged as an important method for assessing and enhancing network security. However, traditional manual penetration testing methods suffer from inefficiency,

收稿日期: 2023-06-12; 修回日期: 2023-10-28

通信作者: 樊成, fancheng_2019@163.com

基金项目: 国家自然科学基金(61902426)

Foundation Item: The National Natural Science Foundation of China(61902426)

引用格式: 樊成, 胡国庆, 丁涛杰, 等. 基于攻击图和深度 Q 学习网络的自动化安全分析与渗透测试模型[J]. 网络与信息安全学报, 2023, 9(6): 166-175.

Citation Format: FAN C, HU G Q, DING T J, et al. Autonomous security analysis and penetration testing model based on attack graph and deep Q-learning network[J]. Chinese Journal of Network and Information Security, 2023, 9(6): 166-175.

human error, and tester skills, leading to high uncertainty and poor evaluation results. To address these challenges, an autonomous security analysis and penetration testing framework called ASAPT was proposed, based on attack graphs and deep Q-learning networks (DQN). The ASAPT framework was consisted of two main components: training data construction and model training. In the training data construction phase, attack graphs were utilized to model the threats in the target network by representing vulnerabilities and possible attacker attack paths as nodes and edges. By integrating the common vulnerability scoring system (CVSS) vulnerability database, a “state-action” transition matrix was constructed, which depicted the attacker’s behavior and transition probabilities in different states. This matrix comprehensively captured the attacker’s capabilities and network security status. To reduce computational complexity, a depth-first search (DFS) algorithm was innovatively applied to simplify the transition matrix, identifying and preserving all attack paths that lead to the final goal for subsequent model training. In the model training phase, a deep reinforcement learning algorithm based on DQN was employed to determine the optimal attack path during penetration testing. The algorithm interacted continuously with the environment, updating the Q-value function to progressively optimize the selection of attack paths. Simulation results demonstrate that ASAPT achieves an accuracy of 84% in identifying the optimal path and exhibits fast convergence speed. Compared to traditional Q-learning, ASAPT demonstrates superior adaptability in dealing with large-scale network environments, which could provide guidance for practical penetration testing.

Keywords: autonomous penetration testing, reinforcement learning, attack graph, deep Q-learning network

0 引言

随着计算机网络的广泛普及以及各类安全事件的不断发生, 信息安全问题逐渐成为关注热点。渗透测试作为有效提升网络安全性的方法, 主要通过专业网络安全人员对目标系统进行授权攻击以发现存在的安全漏洞。然而, 渗透测试主要基于自动化的扫描工具及人工的漏洞利用分析展开, 分析效率较低, 且需要大量人力^[1], 但专业的网络安全人员存在严重不足的问题, 仅我国, 到 2027 年网络安全从业人员缺口将达 327 万^[2], 为此开发基于人工智能的自动化渗透测试工具显得尤为重要。

自动化渗透测试的关键是如何自动寻求目标系统的最优攻击路径。早期的研究主要依赖攻击图对目标进行威胁建模, 通过分析网络漏洞之间的联系来找寻有效攻击路径^[3], 此外还有利用搜索算法遍历攻击图来生成攻击方案的做法^[4]。随着人工智能技术的兴起, 人们逐渐开始思考如何将智能算法应用于渗透测试。一类方法是利用规划区域定义语言 (PDDL, planning domain description language) 将攻击图转换为特定的攻击模型^[5], 并针对不同漏洞生成相应利用规划。Sarraute 等^[6]则在基于 PDDL 的攻击模型的基础上, 对攻击行为添加成功率来描述攻击结果的不确定性, 增加了

该模型对实际渗透测试的适应性。但此类模型难以生成完整的渗透测试方案, 无法满足自动化渗透测试的要求。另一类方法则是将渗透测试看作部分可观测马尔可夫决策过程 (POMDP, partially observable Markov decision process)^[7], 目标网络已知配置即环境状态, 攻击者具体行为即动作, 此时攻击者的攻击行为选择可由 POMDP 的预期奖励来决定。POMDP 的出现为自动化渗透测试提供了新思路, 即利用强化学习算法来寻找最优攻击路径。强化学习算法能够在没有先验知识的前提下, 利用与环境的多次交互生成最优策略, 解决了自动化渗透测试中复杂性和不确定性两大难题^[8]。Yousefi 等^[9]将强化学习与攻击图结合, 使用强化学习对攻击图进行分析, 并利用 Q 学习 (Q-learning) 确定攻击策略。赵海妮等^[10]则提出了无须生成攻击图的自动化 Q 学习渗透测试 (QLPT, Q-learning penetration test) 模型, 该模型通过在环境中多次探索来形成完备的状态-漏洞 Q 表, 使攻击者可根据所处状态选择对应最佳漏洞达到最优路径推荐的效果。但以上方法需要对漏洞利用结果准确建模, 并且受动作空间和样本空间的数量限制, 难以应用于大型网络。

为解决以上问题, 本文提出了基于攻击图与深度 Q 学习网络的自动化安全分析与渗透测试

(ASAPT, autonomous security analysis and penetration testing)模型,通过模拟人工渗透测试的思路,自动化生成对目标网络的最优攻击路径,主要创新点总结如下。

1) 本文所提 ASAPT 模型能够识别网络拓扑及漏洞之间的依赖关系,从而形成对网络在特定域中的“动作-奖励”关系建模。具体来说,模型中生成的转移矩阵与网络拓扑结构以及每个节点通联的复杂性相关。奖励值与每个漏洞的通用漏洞评分系统(CVSS, common vulnerability scoring system)评分相关。因此,ASAPT 能够对现实中的网络渗透过程进行仿真模拟,生成完整的渗透测试方案,并且对实际具有指导意义。

2) 采用基于深度 Q 学习网络(DQN, deep Q-learning network)的强化学习(RL, reinforcement learning)模型对网络中的高效渗透方案进行学习。生成的攻击方案有助于缩短在渗透过程中所耗时间,相对于人工渗透测试具有明显的效率优势。人工的渗透测试所耗时间需要的是“小时/天”,而基于 DQN 的模型是“秒”,更具有优势。

3) 在对训练数据处理的过程中,创新地采用了深度优先搜索(DFS, depth first search)算法对目标网络转移矩阵进行简化。此外,DQN 兼具强化学习与深度学习的优势,能够弥补强化学习在面对大型网络时的不足,以上举措提高了所提模型对于高维大规模网络的处理能力及适应性。

1 背景知识

1.1 漏洞扫描器 Acunetix

Acunetix 是一款进行网络站点分析及漏洞检测的专业扫描软件^[11],支持多线程快速扫描,具备强大的可视化功能,并且对 SQL 注入、跨站脚本攻击(XSS, cross site scripting)等漏洞的检测准确率高。近年来,Acunetix 在 Web 安全领域得到了广泛应用,Vyamajala 等^[12]利用 Acunetix 对特定网站存在的 SQL 漏洞情况进行了分析,并基于分析结果给出了网站维护的相应建议。

1.2 攻击图

攻击图是一种基于模型的网络安全评分技术^[13],它从攻击者角度出发,在综合分析网络配置和漏洞信息的基础上,找出所有可能的攻击路径,并能够可视化攻击场景,从而帮助网络安全管理人

员直观地理解目标网络内各漏洞及配置信息之间的依赖关系以及由此产生的潜在威胁。攻击图广泛应用于网络安全分析,包括区块链网络^[14]、通信网络^[15]、电力网络^[16]、数据中心网络^[17]等。

1.3 强化学习

强化学习是机器学习的一种变体,它允许智能体在特定场景中通过不断感知环境状态与动作奖励来学习调整自身行为,从而自动找到该场景下的最优行动策略,使获得的奖励最大化^[18]。

图 1 展示了强化学习中智能体与环境的交互过程,在时刻 t , 智能体观察到环境状态 s_t 并执行动作 a_t , 此时环境基于该动作产生奖励 r_t , 自身状态随之变为 s_{t+1} 。因此强化学习可以由马尔可夫决策过程(MDP)进行建模,并由五元组 (S, A, P, R, Π) 表示。其中 $S = \{s_1, s_2, \dots, s_n\}$ 为环境的状态空间, $A = \{a_1, a_2, \dots, a_k\}$ 为智能体所采取的动作空间, $P(s' | s, a)$ 表示状态 s 采取动作 a 变为 s' 的转移概率, $R(s, a)$ 表示在状态 s 时动作 a 得到的奖励, Π 表示智能体根据当前状态学习下一个动作所采用的策略集合。

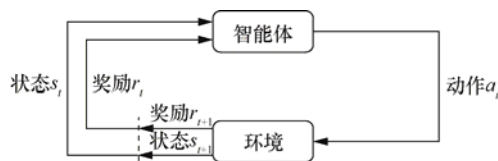


图 1 强化学习中智能体与环境的交互过程
Figure 1 The interaction process between intelligent agents and the environment in reinforcement learning

此时,采用异策时序差分的 Q 学习^[19]算法对智能体所采取的动作集合进行学习,其优化函数为 $Q^*(s, a)$, 表示在状态 s 下采用动作 a 得到的期望奖励,计算式为

$$Q^*(s, a) = R(s, a) + \gamma \sum_{s' \in S} P(s' | s, a') V^*, \quad (1)$$

$$V^* = \max_{a' \in A} Q^*(s', a')$$

其中, V^* 为智能体初始采取最佳行动,并在后续状态下继续选择最佳行动所获得的最优值; γ 表示折扣因子,描述智能体对当前奖励和未来奖励的偏好。通过求解 Q^* , 进而得到最优策略 π^* , $\pi^* = \operatorname{argmax}_{a \in A} Q^*(s, a)$ 。

1.4 DQN

Q-learning 受状态空间和动作空间限制, 难

以应用于大型网络，并且 Q-learning 的智能体无法对未经历的状态和动作做出相应反馈，导致泛化能力不足。为了解决以上问题，文献[20]提出深度 Q 学习网络，该网络将卷积神经网络与 Q-learning 相结合，使用深度神经网络对目标函数 Q 值进行估计，有效解决了 Q-learning 对大型网络适应性不足的问题。DQN 训练过程如图 2 所示。

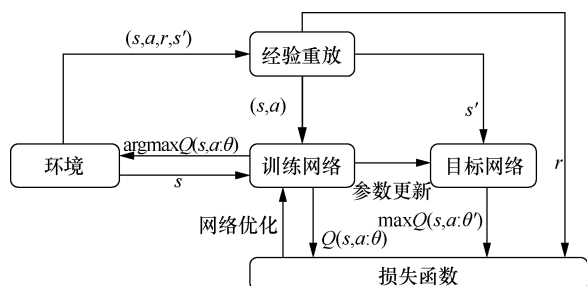


图2 DQN 训练过程
Figure 2 DQN training process

从图 2 可以看出，DQN 从两个角度对 Q-learning 进行了改进。一是采用经验回放的方法^[21]，有效降低了训练样本的数据相关性。具体而言，智能体与环境交互获得的经验 $e_t = \{s_t, a_t, r_t, s_{t+1}\}$ 首先存放在经验集合 $D_t = \{e_1, e_2, \dots, e_t\}$ 中，随后 DQN 从经验集合 D_t 中随机抽取经验样本进行学习。二是通过值近似的方式对 Q 函数进行计算，解决了 Q-learning 在面对大型网络时 Q 值难以计算的问题。DQN 有训练网络和目标网络两个卷积神经网络，训练网络参数 θ 在每次训练中进行更新，目标网络参数 θ' 则在经过一定训练次数后复制 θ 。假设两个网络的输出分别为 $Q(s, a | \theta_i)$ 和 $Q(s, a | \theta'_i)$ ， $Q(s, a | \theta_i)$ 为网络预测值， $R(s, a) + \gamma \max Q(s, a | \theta'_i)$ 为网络实际值，那么网络损失函数可以表示为

$$L(\theta_i) = E_{s,a,r,s'} [(R(s,a) + \gamma \max Q(s,a|\theta'_i) - Q(s,a|\theta_i))^2] \quad (2)$$

再依据随机梯度下降法对训练网络进行参数更新。

2 模型总体架构

本文提出的 ASAPT 模型架构如图 3 所示，主要由两部分组成，分别是训练数据构建以及

DQN 训练数据构建。

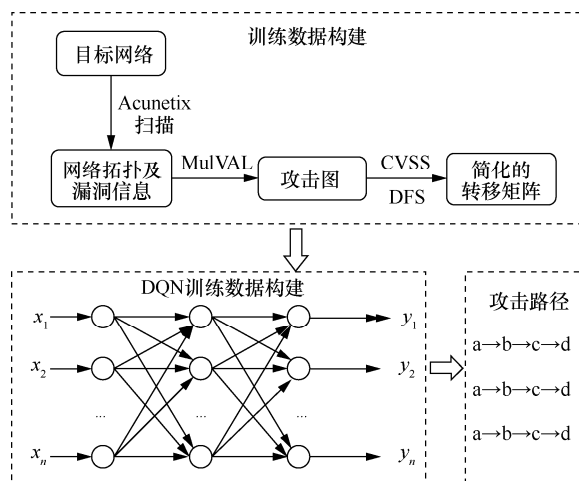


图3 ASAPT 模型架构
Figure 3 Architecture of ASAPT

(1) 训练数据构建

使用 Acunetix 对目标网络进行扫描，收集网络拓扑及漏洞信息。随后结合 CVSS 对各漏洞的评分，构建目标网络信息库，将该信息库输入 MulVAL，生成对目标网络的攻击图。

图 4 以 Web 服务器节点为例，对目标网络信息库进行了展示，可以看出，该库包含 Web 服务器节点的名称 (webserver)、协议 (httpd)、开放的端口 (80) 以及运行的应用 (apache)，存在有漏洞 (CVE-2021-42013)，及漏洞利用效果 (privEscalation)。

```
[vulExists(webserver, CVE-2021-42013, httpd)
vulProperty(CVE-2021-42013, remote-Exploit,
privEscalation)
networkServiceInfo(webserver, httpd,tcp,80,apache)]
```

图4 目标网络信息库示例

Figure 4 Example template of target network information base

(2) DQN 训练数据构建

对上一步骤生成的攻击图进行处理，形成相应的转移矩阵，并经简化后作为 DQN 的输入。在处理过程中，需要对攻击图各节点赋予对应奖励值。对于漏洞利用的节点，本文采用基于 CVSS 库的漏洞评分[该评分包含基础性得分 (base score) 和便利性得分 (exploitability score) 两项]，依据特定的公式计算得到^[22]；对于其他节点，则采用如下的赋值规则。

1) 初始节点的奖励值设为 0.01, 最终目标节点的奖励值设为 100。

2) 对于能够实现代码执行或文件访问的节点, 也可以获取奖励值, 本文设置为 1.5, 这些行为对实际的渗透测试有较大帮助。

3) 对于任意两个没有路径连接的节点, 其得分分为-1, 其余的节点得分为 0。

据此, 将攻击图的各节点及其奖励值映射到矩阵中, 形成转移矩阵。此时的转移矩阵包含渗透测试中所有可能的攻击动作, 但并非所有的攻击动作以及其形成的路径都能达到最终的攻击目标, 这些无法到达最终目标的路径对于 DQN 训练毫无益处。因此需要挑选出转移矩阵中所有能到达最终目标的路径所包含的节点及动作, 形成简化版的转移矩阵。此处利用深度优先搜索算法对转移矩阵进行处理, 找到所有能够达到最终攻击目标的路径, 结合各路径的奖励分数, 形成简化版的转移矩阵, 并将此作为 DQN 的训练数据。

DQN 由训练网络和目标网络构成, 两个网络初始时刻配置参数完全相同, 均为由 3 个卷积层及 2 个全连接层构成的卷积神经网络, 各卷积层的激活函数为 RELU。在训练过程中, 其输入为简化版转移矩阵, 经由智能体与环境的交互得到经验集合, 随后在其中抽取固定长度的经验对输入训练网络和目标网络, 根据两个网络的输出及式(2), 采用随机梯度下降法对训练网络进行参数更新。目标网络则在固定步长后复制训练网络参数。

训练结束后, 即可将训练好的目标网络用于实际网络的最优渗透路径确定。具体而言, 将待攻击网络的简化版转移矩阵输入 DQN 的目标网络, 得到起始状态下不同动作对应的 Q 值, 选择使 Q 值最大的动作作为该状态下的路径, 进而得到完整的攻击路径。

3 实验与结果分析

3.1 实验场景搭建

为验证本文模型, 此处搭建实验场景进行验证, 实验场景的网络拓扑结构如图 5 所示。

该场景表示一个小型的公司网络, 包括一台 Web 服务器、一台文件服务器和一台工作站主机。文件服务器和工作站主机处于同一子网络, 并经路由器与防火墙相连接, Web 服务器则处于另一

子网络, 并经防火墙连接至外部互联网。防火墙为外部防火墙, 且具有以下规则。

1) 互联网端的攻击者能够通过 HTTP (hypertext transfer protocol) 及相应端口访问该网络 Web 服务器。

2) Web 服务器能够通过 NFS (network file system) 协议、FTP (file transfer protocol) 等与另一子网中的文件服务器相互通联。

3) 文件服务器可以通过 NFS 协议、Samba 协议等与工作站主机相互通联。

4) 文件服务器及工作站主机能够通过 HTTP 访问互联网。

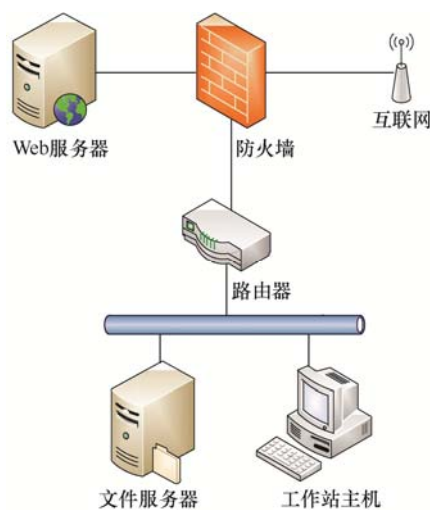


图5 实验场景的网络拓扑结构

Figure 5 The topology structure of the experiment environment

3.2 实验步骤

本节对 ASAPT 模型的性能进行仿真分析。其中, 网络扫描部分使用标准漏洞扫描软件 Acunetix, 攻击图则由 MulVAL 工具生成, 同时基于网络配置、漏洞信息以及对应 CVSS 评分来建立状态-动作转移矩阵, 并利用 DFS 算法对上述矩阵进行简化, 形成简化后的奖励矩阵作为 DQN 的训练数据。

(1) 目标网络信息搜集

使用 Acunetix 对场景内的 Web 服务器及文件服务器进行扫描, 对其上存在的开放端口、运行协议及可能存在的漏洞信息进行确认, 得到目标网络配置信息 (如表 1 所示)。再依据 CVSS 对存在漏洞的描述, 包括类型、效果、等级评分等, 得到目标网络漏洞信息, 如表 2 所示。

表 1 目标网络配置信息

Table 1 Target network configuration information

主机	漏洞标识	应用	端口	对应服务
Web 服务器	CVE-2021-42013	Apach	80	HTTP/HTTPS
文件服务器	CVE-2019-12815	FTP	21	FTP
工作站	—	—	—	—

表 2 目标网络漏洞信息

Table 2 Target network vulnerability information

漏洞标识	漏洞类型	漏洞得分	效果
CVE-2021-42013	代码执行	9.8	root
CVE-2019-12815	文件读取	9.3	user

从表 1 和表 2 可以看出，网络拓扑中的软件漏洞共有 2 种，Web 服务器具有标识为“CVE-2021-42013”的漏洞，该漏洞能够使攻击者通过路径穿越访问服务器目录，并远程执行代码。文件服务器具有标识为“CVE-2019-12815”的漏洞，该漏洞与 ProFTPD 模块相关，能够使攻击者任意浏览文件，造成信息泄露。工作站主机只运行简单的传输协议，不存在可利用漏洞。

(2) 训练数据构建

将上一步骤中搜集的目标网络环境的拓扑结构及配置信息库输入 MulVAL，生成实验场景对应的攻击图，如图 6 所示。

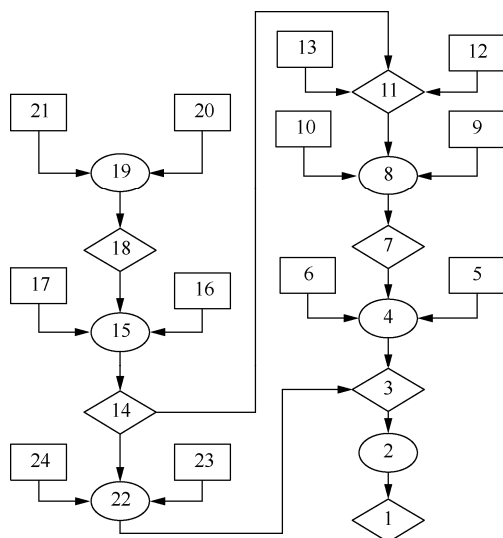


图 6 实验场景对应的攻击图

Figure 6 Attack graph of test environment

图 6 展示了攻击者对目标网络的攻击路径，每个节点都对应配置信息及漏洞利用方式，攻

击图节点信息描述如表 3 所示。在对目标网络的攻击过程中，攻击者利用互联网从外部进入，因此起点为“节点 21”，其位置在图左上方。最终目标是能够在工作站主机上执行代码，因此终点为“节点 1”，其位置在图右下方。攻击者可以依据图中箭头方向进行移动，直到到达最终的攻击目标。此外，图中方形节点（如节点 21、17、13）表示系统的配置信息；菱形节点表示攻击者能够获取的潜在权限或渗透路径（如在 Web 服务器执行代码，提升至 root 权限）；椭圆节点连接攻击前后的状态，表示攻击者根据目标系统信息进行漏洞利用所取得的结果（如对节点 15 来说，攻击者通过漏洞利用，达到了在节点 14 进行执行代码的结果）。

表 3 攻击图节点信息描述

Table 3 Description of the attack graph nodes information

序号	节点信息描述
1	execCode(workStation, root), OR, 0
2	RULE 4(Trojan horse installation), AND, 0
3	accessFile(workStation, write, 'export'), OR, 0
4	RULE10(NFS shell), AND, 0
5	nfsExportInfo (workStation, 'export', write,fileServer), LEAF, 1
6	hacl(fileServer, workStation, nfsProtocol, nfsPort), LEAF, 1
7	execCode(fileServer, user), OR, 0
8	RULE3(remote exploit of a client program), AND, 0
9	vulExists(fileServer, 'CVE-2019-12815', ftp, remote-Exploit, privEscalation), LEAF, 1
10	networkServiceInfo(fileServer, http, tcp ,80, root), LEAF, 1
11	netAccess(fileServer, tcp ,80), OR 0
12	RULE7(multi-hop access), AND, 1
13	hacl(webServer, fileServer, tcp, 80), LEAF, 1
14	execCode(webServer, root), OR, 0
15	RULE2(remote exploit of a server program), AND, 0
16	vulExists(webServer, 'CVE-2021-42013', remote-Exploit, privEscalation), LEAF, 1
17	networkServiceInfo(webServer, httpd, tcp ,80),LEAF, 1
18	netAccess(webServer, tcp. 80), OR, 0
19	RULE6(direct network access), AND, 0
20	hacl(internet, webServer, tcp, 80), LEAF, 1
21	attackerLocated(internet), LEAF, 1
22	RULE15(NFSshell), AND, 0
23	nfsExportInfo(fileServer, 'export', write, webServer), LEAF, 1
24	hacl(webServer, fileServer, nfsProtocol, nfsPort), LEAF, 1

依据表 3, 结合各节点的奖励值 (依据式(3)计算), 形成对目标网络的状态-动作转移矩阵, 攻击图对应转移矩阵如图 7 所示, 其中矩阵的行表示当前状态下所采取的动作, 矩阵的列表示攻击者所处状态。如 $R(21, 16)$ 表示攻击者从节点 21 进入节点 16, 并获取了 9.8 的奖励; $R(9, 23)$ 值为 -1, 表示这两个节点之间没有直接相连, 无法转移; $R(5, 7)$ 值为 0, 表示攻击者可由节点 5 返回节点 7。

状态	R	动作								
		21	16	14	23	3	9	7	5	1
状态	21	0.01	9.8	-1	-1	-1	-1	-1	-1	-1
	16	0	-1	1.5	-1	-1	-1	-1	-1	-1
	14	-1	0	-1	0	-1	9.3	-1	-1	-1
	23	-1	-1	0	-1	1.5	-1	-1	-1	-1
	3	-1	-1	-1	0	-1	-1	-1	-1	100
	9	-1	-1	0	-1	-1	-1	1.5	-1	-1
	7	-1	-1	-1	-1	-1	0	-1	1.5	-1
	5	-1	-1	-1	-1	-1	-1	0	-1	100
	1	-1	-1	-1	-1	0	-1	-1	0	100

图 7 攻击图对应转移矩阵

Figure 7 The transition matrix of the attack graph

但图 7 含有无法抵达最终目标的无效路线, 若直接作为 DQN 的训练数据, 会影响模型的训练效率, 不利于模型在大规模网络中应用, 因此需要对图 7 中的数据进行简化。本文使用 DFS 算法对上述转移矩阵进行简化, 去除无法抵达最终目标的路径, 仅保留能够到达最终目标的路径, 并对保留路径中的中间节点进行合并, 就能得到可行路线及奖励值, 结果表 4 所示。其中第 2 列为各路径起始节点的奖励值, 第 3 列为各路径中间节点的累积奖励值, 最后一列为最终目标节点的奖励值。

表 4 利用 DFS 算法得到的可行路线及奖励值
Table 4 Feasible routes and their rewards obtained using the DFS algorithm

攻击路径 编号	起始节点 的奖励值	中间节点的 累积奖励值	最终目标节点 的奖励值
1	0.01	12.8	100
2	0.01	22.1	100

简化后的转移矩阵如图 8 所示, 将此作为 DQN 的训练数据。

R	状态	动作			
		0	1	2	3
0		0.01	12.9	22.1	-1
1		0	-1	-1	100
2		0	-1	-1	100
3		-1	0	0	-1

图 8 简化后的转移矩阵

Figure 8 Simplified transition matrix

(3) DQN 训练

DQN 基于 Pytorch 深度学习框架及 gym 工具包实现。操作系统为 Windows 10 (64 bit), 内存为 16 GB, 处理器为 Intel Core i7-6500U2, 59 GHz, 显卡为 NVIDIA GeForce RTX 3080ti。此外, 为满足 DQN 训练需求, 需要增大训练数据数量。这里依据目标网络的配置信息, 从漏洞信息库中选择对应的漏洞赋予相应的 Web 服务器及文件服务器, 形成不同的漏洞组合, 每种组合生成一个攻击图, 共得到 300 个攻击图。随后随机抽取其中 200 个攻击图作为训练集, 剩余 100 个攻击图作为验证集。对于验证集数据而言, 将步骤最少且达到攻击目标后奖励值最大的路径作为最优路径。网络的最大迭代次数设置为 1 000, 而后开始训练, 训练结束后, 在验证集数据上进行验证, 并生成对目标网络的最优攻击路径。

3.3 实验结果分析

(1) 训练结果分析

图 9 展示了本文提出的 ASAPT 模型在实验场景 100 次训练过程中, 其攻击路径的最大奖励值变化曲线。可以看出, 在训练初期, 获得的最大奖励值较小, 在经过 20 次迭代训练后, 奖励值不断升高, 在经过 70 次迭代训练后, 奖励值逐渐稳定, 说明 ASAPT 模型已经找到最优的攻击路径, 结果为“16→8→11→13→9→5→7→1”。结合图 6 中攻击图对该结果进行分析, 可以看出该路径充分利用目标网络中 Web 服务器和文件服务器间的通连路径, 采用的攻击方法易于实现, 能够达到最优攻击效果。

随后在训练集对 APAPT 模型进行评估, ASAPT 模型性能表现如表 5 所示。从训练后的

APAPT 模型在验证集数据中的表现可以看出, 本文所提 APAPT 模型在该实验场景下表现良好, 能够以较高的准确率及较短的路径长度实现对目标网络的渗透。

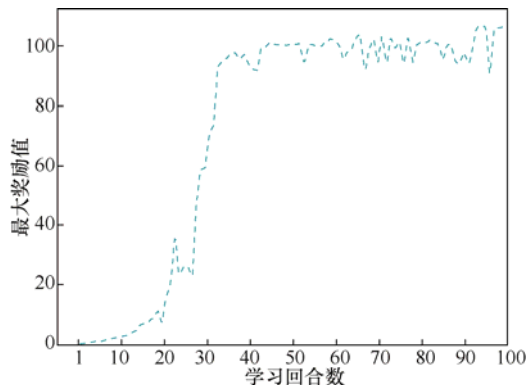


图9 攻击路径的最大奖励值变化曲线

Figure 9 The maximum reward changes curve for attack path

表5 ASAPT 模型性能表现
Table 5 The performance of ASAPT model

模型	准确率	平均路径长度
ASAPT	0.837	1.925

(2) DQN 参数分析

为了进一步研究 DQN 参数对训练结果的影响, 此处选择批容量 (BS, batch size) 和折扣系数 γ 这两个参数, 分析其在不同取值下, 模型收敛时间和最大奖励值的变化。其中, γ 的变化范围为[0.6,0.99], BS 则选择 16、32 及 64。运行时间随不同的折扣因子 γ 和批容量的变化情况如图 10 所示。

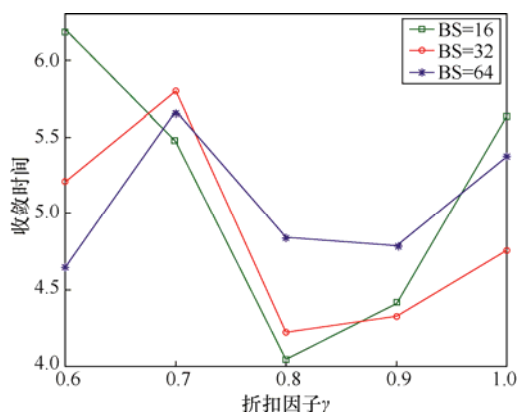


图10 收敛时间随不同的折扣因子 γ 和批容量的变化情况
Figure 10 The convergence time of ASAPT framework obtained by variation of discount γ and batch size

从图 10 可以看出, DQN 在不同的 γ 下, 均实现了较快收敛。当 $\gamma=\{0.6,0.7\}$ 时, 收敛时间在 4.5 s 和 6.5 s 之间变化; 当 $\gamma=0.8$ 时, 收敛速度最快, 说明对未来状态较高的认可度和探索概率对 DQN 的收敛有积极作用; 当 $\gamma=\{0.9,0.99\}$ 时, DQN 相较于 $\gamma=0.8$ 时更偏向于未来状态, 而对当前状态的认可度降低, 此时 DQN 收敛速度变慢, 说明不能过于忽视当前状态而一味增加对未来状态的重要程度。此外, 当 $\gamma=\{0.7,0.8\}$ 时, 与 BS= {32,64} 相比, BS=16 时 DQN 的收敛速度更快, 说明 BS=16 对小规模网络目标的渗透更为合适。

图 11 展示了最大奖励值随不同的折扣因子 γ 和批容量的变化情况。当 $\gamma=0.8$ 时, DQN 取得了最大奖励值, 随着 γ 的增加, DQN 的最大奖励值逐渐减少。此外, 当 BS=16 时, DQN 的奖励值高于 BS 的其他取值。因此, $\gamma=0.8$, BS=16 对于该小规模实验场景是最佳参数设置取值, 此时 DQN 收敛速度快, 取得的最大奖励值高。对于其他大规模网络, 如物联网和大型互联网, 可以考虑增大 BS 来提升模型性能。

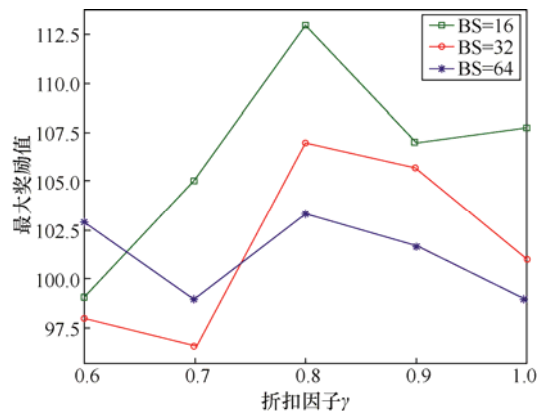


图11 最大奖励值随不同的折扣因子 γ 和批容量的变化情况
Figure 11 The maximum reward of ASAPT framework obtained by variation of discount γ and batch sizes

4 结束语

本文提出基于攻击图和深度Q网络的自动化渗透测试模型, 该模型首先使用 Acunetix 扫描软件搜集目标网络的主机和漏洞数据, 以构建真实的训练和验证场景; 随后使用攻击图为每个训练场景生成攻击信息, 并结合 CVSS 对每个

节点赋予奖励分数,而后对这些信息进行处理并用于训练 DQN;最后使用训练好的 DQN 来确定目标网络场景的最优攻击路径。本文搭建了实验场景(在给定的网络拓扑中赋予随机的漏洞数据,以构建 200 个训练和 100 个验证场景)对模型的性能进行验证,结果表明所提模型识别最优攻击路径方面准确率高,收敛速度快,能够在实际渗透测试过程中给出合理的攻击策略。

下一步工作中,一是用更多实际网络场景来扩展训练数据集,以提高 ASAPT 模型的通用性和稳定性;二是在所提模型中集成网络服务扫描功能,以便将真实目标环境的信息自动提供给 ASAPT 模型,从而获得更准确的实际网络拓扑结果。

参考文献:

- [1] JAJODIA S, NOELS. Topological vulnerability analysis: a powerful new approach for network attack prevention, detection, and response[M]//Algorithms, Architectures and Information Systems Security. World Scientific, 2008: 285-305.
- [2] 网络安全人才实战能力白皮书[EB]. 2022.
White paper on practical competence of cybersecurity talents[EB]. 2022.
- [3] WANG G Y, WANG H M, CHEN Z J, et al. Research on computer network attack modeling based on attack graph[J]. Journal of National University of Defense Technology, 2009, 31(4): 74-80.
- [4] SHEYNERO, HAINESI, JHA S, et al. Automated generation and analysis of attack graphs[C]//Proceedings 2002 IEEE Symposium on Security and Privacy. 2004: 273-284.
- [5] OBESJL, SARRAUTEC, RICHARTE G. Attack planning in the real world[EB].
- [6] SARRAUTE C, RICHARTE G, OBES J L. An algorithm to find optimal attack paths in nondeterministic scenarios[C]//Proceedings of the 4th ACM Workshop on Security and Artificial Intelligence. 2011: 71-80.
- [7] SARRAUTE C, BUFFET O, HOFFMANN J. Penetration testing=POMDP solving[J]. Computer Science, 2013, 33(5): 535-540.
- [8] JONATHON S, HANNA KI. Autonomous penetration testing using reinforcement learning[J]. arxiv preprint arxiv: 1905. 05965, 2019.
- [9] YOUSEFI M, NHAMO M, ZHANGY, et al. A reinforcement learning approach for attack graph analysis[C]//17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE). 2018. 212-217.
- [10] 赵海妮, 焦健. 基于强化学习的渗透路径推荐模型[J]. 计算机应用, 2022, 42(6): 1689-1694.
ZHAO H N, JIAO J. Recommendation model of penetration path based on reinforcement learning[J]. Journal of Computer Application, 2022, 42(6): 1689-1694.
- [11] SECURITY A. Acunetix Web application security blog: statistics from 10, 000 leaked hotmail passwords[EB].
- [12] VYAMAJALA S, MOHD T K, JAVAID A. A real-world Implementation of SQL injection attack using open source tools for enhanced cybersecurity learning[C]//Proceedings of 2018 IEEE International Conference on Electro/Information Technology(EIT). 2018.
- [13] 康海燕, 龙墨澜. 基于吸收马尔可夫链攻击图的网络攻击分析方法研究[J]. 通信学报, 2023, 44(2): 122-135.
KANG H Y, LONG M L. Research on network attack analysis method based on attack graph of absorbing Markov chain[J]. Journal on Communications, 2023, 44(2): 122-135.
- [14] 罗智勇, 宋伟伟, 张文博, 等. 基于 Markov 攻击图和博弈模型的区块链安全态势感知方法[J]. 电子与信息学报, 2023, 45(4): 1374-1382.
LUO Z Y, SONG W W, ZHANG W B, et al. Blockchain security situational awareness method based on Markov attack graph and game model[J]. Journal of Electronics & Information Technology, 2023, 45(4): 1374-1382.
- [15] 王赛娥, 刘彩霞, 刘树新. 一种基于攻击图的 5G 网络安全风险评估方法[J]. 计算机应用与软件, 2023, 40(4): 289-296, 335.
WANG S E, LIU C X, LIU S X. A method of 5G network security risk assessment based on attack graph[J]. Computer Applications and Software, 2023, 40(4): 289-296, 335.
- [16] KANG J, ZHANG Y, ZHOU Z. A bayesian attack graph-based vulnerability assessment for cyber-physical power systems[J]. IEEE Transactions on Smart Grid, 2023, 14(2): 139-148.
- [17] SABUR A, CHOWDHARY A, HUANG D, et al. S3: a DFW-based scalable security state analysis framework for large-scale data center networks[C]//The 22nd International Symposium on Research in Attacks, Intrusions and Defenses(RAID). 2019: 497-485.
- [18] JEFFREY D J, LI J H, CHEN Z S. Reinforcement learning: an introduction[J]. Neurocomputing, 2000, 35(1).
- [19] CHRISTOPHER J. Q-learning[J]. Machine Learning, 1992, 8(3), 279-292.

- [20] FAN J Q, WANG Z R, XIE Y C, et al. A theoretical analysis of deep q learning[C]//Proceedings of the 2nd Conference on Learning for Dynamics and Control/Proceedings of Machine Learning Research (PMLR). 2020: 486-489.
- [21] WANG Y, WANG Y, LI Y, et al. A review on deep Q-learning for robotic control[J]. IEEE Transactions on Industrial Electronics, 2022, 69(11): 9468-9481.
- [22] 潘刚, 米士超, 郭荣华, 等. 基于攻击树和 CVSS 的网络攻击效果评估方法[J]. 电子技术应用, 2022, 48(4): 76-80.
- PAN G, MI S C, GUO R H, et al. Evaluation method of network attack effect based on attack tree and CVSS[J]. Application of Electronic Technique, 2022, 48(4): 76-80.



胡国庆（1982- ），男，四川泸州人，中国人民解放军 31681 部队助理研究员，主要研究方向为电子信息、网络安全。



丁涛杰（1990- ），男，浙江诸暨人，中国人民解放军 31681 部队研究实习员，主要研究方向为普适计算、网络安全。

[作者简介]



樊成（1993- ），男，甘肃庆阳人，中国人民解放军 31681 部队研究实习员，主要研究方向为网络信息系统安全、数据挖掘和人工智能安全。



张展华（2000- ），男，河北磁县人，中国人民解放军 31681 部队研究实习员，主要研究方向为网电对抗、人工智能安全。