基于强化学习的自免疫动态攻击生成方法

李 腾,唐智亮,马 卓,马建峰
(西安电子科技大学网络与信息安全学院,陕西西安 710071)

摘 要: 通过最优路径发动网络攻击的方式已成为影响各企业、机构内部网络安全的重要因素. 现有的针对内部网络探寻最优攻击路径大多是采用攻击图的方式实现,未考虑攻击代价和攻击利益的关系,已有的利用*Q*-learning算法机制解决攻击路径的方法存在网络脆弱性信息利用率低的问题. 为解决这些问题,本文借鉴生物免疫机制提出了一种基于强化学习的自免疫动态攻击生成方法,模拟攻击者对内网的网络攻击,从而高效地发现内部网络中存在的脆弱点,实现自免疫防御. 方案首先对内部网络信息进行窃取并加以处理,在攻击图的有向边上附加权值,然后通过改进的*Q*-learning算法寻找最优攻击路径,实现全部最优攻击路径的获取,并返回最优攻击路径的攻击图和内部网络主机脆弱性分析结果. 通过理论分析和实验结果表明,该方法兼顾寻找最优攻击路径的高效性、准确性的同时,还解决了最优攻击路径中存在环型回路、多条最优攻击路径的问题,充分利用内部网络脆弱性,提升自免疫安全防护能力.

关键词: 最优攻击路径;强化学习;攻击图;路径规划;内网安全
基金项目: 国家自然科学基金(No.62272370);中国科协青年人才托举工程(No.2022QNRC001);陕西省科学技术协会青年人才托举计划(No.20210120)
中图分类号: TP309.1 **文献标识码:** A **文章编号:** 0372-2112(2023)11-3033-09
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20230369

Autoimmune Dynamic Attack Generation Method Based on Reinforcement Learning

LI Teng, TANG Zhi-liang, MA Zhuo, MA Jian-feng
(*School of Cyber Engineering*, *Xidian University*, *Xi'an*, *Shaanxi* 710071, *China*)

Abstract: The approach of launching network attacks through optimal pathways has become a significant factor affecting the internal network security of various enterprises and organizations. Existing methods for exploring optimal attack pathways within internal networks mostly rely on attack graphs and often neglect the relationship between attack costs and benefits. Methods that utilize the *Q*-learning algorithm to address attack pathways suffer from low utilization of network vulnerability information. To address these issues, this paper draws inspiration from the biological immune system and proposes a reinforcement learning-based dynamic self-immune attack generation method. This method simulates network attacks by intruders on an internal network, efficiently uncovering vulnerabilities within the internal network, thereby achieving self-immune defense. The proposed approach first acquires and processes internal network information, attaches weights to directed edges in the attack graph, and then employs an improved *Q*-learning algorithm to discover optimal attack pathways. It successfully identifies all optimal attack pathways, providing attack graphs and an analysis of host vulnerabilities within these pathways. Theoretical analysis and experimental results demonstrate that this method not only efficiently and accurately identifies optimal attack pathways but also resolves issues such as ring loops and multiple optimal attack pathways. By making full use of internal network vulnerabilities, it enhances self-immune security defenses.

Key words: optimal attack path; reinforcement learning; attack graph; path planning; intranet security
Foundation Item(s): National Natural Science Foundation of China (No.62272370); Young Elite Scientists Sponsorship Program by CAST (No.2022QNRC001); Shaanxi University Science and Technology Association Youth Talent Promotion Project (No.20210120)

收稿日期:2023-04-21;修回日期:2023-09-27;责任编辑:覃怀银

1 引言

随着网络系统日益复杂化、多元化,网络风险也在不断上升. 传统的漏洞扫描技术已经不能全面评估网络漏洞所带来的潜在安全问题,攻击者在掌握漏洞之间的关联关系后以跳板攻击的形式对网络发起攻击,经过多步攻击后达到攻击目标主机的目的. 2022 年 6 月,西北工业大学发布《公开声明》称,该校遭受境外网络攻击,经过技术团队分析,判定源头来源于美国国家安全局特定入侵行动办公室(office of Tailored Access Operation,TAO). 此次网络攻击行动中,TAO 选择了中国周边国家的教育机构、商业公司等网络应用流量较多的服务器为攻击目标,控制了大批跳板机,从而窃取了大量关键敏感信息. 由于此类跳板攻击具有攻击性高、难以溯源的特点,传统的安全工具难以发现和应对潜在脆弱点及其组合带来的安全风险,因此迫切需要从攻击者的角度模拟基于最优攻击路径方式的网络攻击,发现内部网络中潜在的脆弱点,从而加强内网安全防御.

为解决这一问题,大量学者对内部网络中探寻最优攻击路径展开了研究. 当下主流方案是采用攻击图的方式去实现[1~3],该方式是一种基于模型的网络脆弱性分析技术,利用攻击图可以发现潜在的攻击路径. 攻击图可分状态攻击图和属性攻击图,状态攻击图是由法国科学家 Dacier 等人首次提出[4],可以适用于小型内部网络分析. 属性攻击图[5]能适应不同规模网络环境,但是会存在环型攻击路径问题. 乔治梅森大学的 Ramakrishnan 和 Ritchey 最初研究攻击路径的时候就使用模型检测方法生成攻击图去描述多步攻击路径[6,7]. 网络安全研究人员 Ritchey 和 Ammann 利用模型检测器 SMV 去构造一个异构网络的攻击图[7],但每一次只能得出一条攻击实例.

攻击路径规划技术是人工智能技术在网络空间安全领域中的重要应用,其方法大多采用前向或后向搜索寻找规划解[8],李庆朋等人提出基于更改智能算法 ACO(Ant Colony Optimization)的最优攻击路径寻找方式[9],但未考虑攻击代价和攻击利益的关系;杨本毅在基于攻击图的渗透方式中通过 Dijkstra 算法去寻找最短攻击路径[10],并且通过赋予权值的方式引入各个路径相关参数,但是并未考虑到网络环境中含有环型回路等复杂情况.

强化学习技术是在未知网络环境进行最优路径探寻的主要方式[11]. 目前在内部网络中探寻最优攻击路径往往是动态的、非确定性、部分观测条件下的路径发现,这导致需要研究强化学习技术这类不确定性规划技术[12,13],帮助攻击者在环境知识不完整的条件下进行路径规划,其中,Q-learning 算法是一种无模型、在线学习的强化学习算法,通过训练能快速寻找最短路径[14,15],同时可以适应未知环境,因此十分适合做最优路径规划工作. 已有的利用 Q-learning 算法机制解决最优攻击路径的方法解决了空间复杂度高的问题[16~18],但是对于其方法运用缺乏针对性,占用了较多内存资源.

因此,本文提出了一种基于强化学习的自免疫动态攻击生成方法,较为系统性地给出内网最优攻击路径寻找过程,利用网络环境和生物系统运行机制相似的特性,实现类免疫防御.

2 背景知识

强化学习技术通过与环境的交互进行学习,最终生成动作策略,可以在任何给定状态下采取最大化长期收益的动作. 其中 Q-learning 算法是一种 Value-Based 算法[19]. Q 所指状态动作价值函数 $Q(s,a)$,表示在状态集合 $\varphi_s\{s_1,s_2,s_3,\cdots\}$ 的某一种状态 s_n 下,采取的动作集合 $\varphi_a\{a_1,a_2,a_3,\cdots\}$ 中某一行动 a_n 所带来的收益期望,收益期望被保存在 Q-table 中,如表 1 所示.

表 1 Q-learning 算法中 Q-table

Q-table	a_1	a_2	a_3
s_1	$Q(s_1,a_1)$	$Q(s_1,a_2)$	$Q(s_1,a_3)$
s_2	$Q(s_2,a_1)$	$Q(s_2,a_2)$	$Q(s_2,a_3)$
s_3	$Q(s_3,a_1)$	$Q(s_3,a_2)$	$Q(s_3,a_3)$
…	…	…	…

表 1 中每一个期望值满足 $Q(s\in\varphi_s,a\in\varphi_a)$. 期望值的初始数据通常是一致的,之后通过马尔科夫抉择过程,在某一状态 State 下,依据当前期望值 $Q(s_n,a_n)$ 选择动作 Action 执行并从环境中获取选择该动作的奖励 Reward,后由状态价值函数处理这些数据并更新该状态下选择动作 Action 的期望值 $Q'(s_n,a_n)$. 经历多次这样选择动作、获取环境奖励、更新期望值的过程后,Q-table 所代表的智能体就成功完成经验积累.

Q-learning 算法的主要优势在于,融合了动态规划和蒙特卡洛的时间差分法(Temporal Difference,TD),并通过 ε-贪婪算法进行平衡探索和利用. Q-learning 算法的价值函数 $Q(s,a)$ 的更新如式(1)所示:

$$Q(s,a)\leftarrow Q(s,a)+\alpha\left(R+\gamma\max_{a'}Q(s',a')-Q(s,a)\right) \quad (1)$$

其中,α 表示学习率,决定每次期望值和实际奖励值的误差值被学习的程度;R 表示状态 s 下选取的动作 a 得到的奖励值;$\gamma\in[0,1]$ 表示折扣因子,为对未来奖励的衰减值;s' 表示在状态 s 下选择动作 a 后的更新状态;$\max_{a'}Q(s',a')$ 表示在新的状态下,选择动作 a 时可以达到当前状态下的最大期望值.

3　系统概览

基于强化学习的自免疫动态攻击路径寻找方案，总体设计框图如图1所示．方案的设计大致可以分为网络拓扑发现与构建、端口扫描与漏洞扫描、攻击图构建与攻击路径获取三个部分．网络拓扑发现和构建模块包含内网网段发现、主机存活性扫描、网络拓扑图构建功能．端口扫描和漏洞扫描模块是在已知网络内部主机存活的前提下，完成主机开放端口和端口漏洞发现任务，并根据得到的漏洞信息计算攻击图权值．攻击图构建与攻击路径获取实现了含权值的有向边攻击图构建、改进 Q-learning 算法实现、最优攻击路径获取以及多条攻击路径绘制与汇总功能，从而获取内网关键脆弱点加强节点防御．

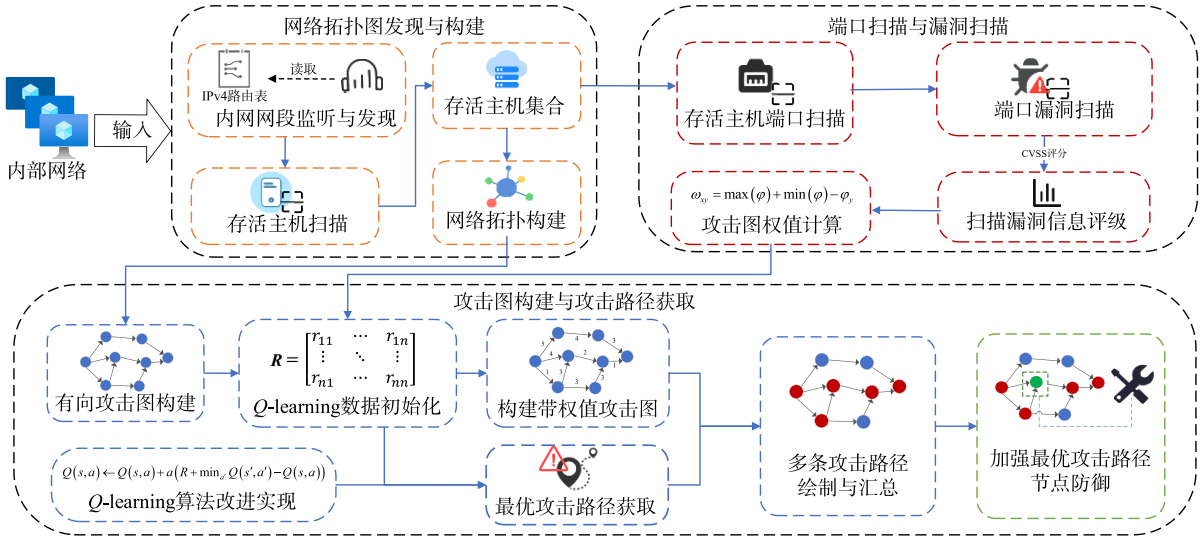


图1　基于强化学习的自免疫动态攻击生成方法总体设计

4　系统设计

4.1　内部网络信息获取

4.1.1　网段发现和主机存活性检测

内网信息获取是内网渗透的必要过程，通过对路由器或者某些主机的路由表信息读取，然后加以信息加工，便可以获得与该设备相连接的网段信息．本文通过读取某主机的路由表中IPv4活动路由表信息，然后进行信息处理，并记录可能存在的网段．

本文采用基于Nmap扫描方式进行网络主机存活性检测．Nmap扫描方式可以适应很多未知的网络环境，通过调节Nmap参数，可以获取多方面的内网信息．本文使用Nmap的Ping扫描方式(参数为-sn)，且最大并发端口数量为100.

4.1.2　基于ICMP协议的网络拓扑发现

基于ICMP协议[20,21]的网络拓扑发现方式所用的探测工具一般有两种，Ping指令和Traceroute指令．从Pansiot和Grad于1998年发表的关于Traceroute网络拓扑测量论文[22]开始，陆续出现了相关方面的研究[23]，本文使用Traceroute指令完成路由跟踪，Ping指令构建存活主机列表，首先对网段内存活的IP地址进行发现和记录，然后对已知存活IP实现目标跟踪，记录相邻两个IP地址的情况，并用图论的方式将其记录下来．

4.1.3　端口扫描与漏洞量化

端口扫描是对主机漏洞扫描的铺垫，不同的主机由于启用服务不同，开放的端口也不同．本文将漏洞风险作为攻破主机代价的评估．漏洞量化是将漏洞风险进行评级，综合多方面信息对每一个漏洞进行合理的评估，并整理参与攻击图的权值计算．

由于目标主机可能存在多个开放端口，所以采用多线程方式对已知目标主机的多个开放端口同时进行漏洞扫描，可以节约大量漏洞发现时间．漏洞扫描首先从主机-端口集合里面依次取出存活主机的开放端口并创建新线程，然后以端口号去匹配漏洞库，验证漏洞是否存在，如果存在则保存数据．本文采用美国国家漏洞数据库提供的CVSS(Common Vulnerability Scoring System)工具进行实验．

4.2　Q-learning算法实现

4.2.1　Q-learning算法奖励矩阵计算

奖励矩阵的值由设备漏洞评分决定．根据漏洞扫描结果可以得到漏洞评分，漏洞评分越高危险程度越大，攻击者就能使用较少的时间和资源攻击目标主机拿到高级权限，在攻击图中本次攻击行为所指向的边的权值就较小．

总结漏洞评分和权值关系为，漏洞评分越高，权值

越小;漏洞评分越低,权值越大. 漏洞评分到权值的如式(2)所示:

$$\omega_{xy}=\max(\varphi)+\min(\varphi)-\varphi_y \tag{2}$$

其中,x表示源节点,y表示指向节点,φ表示漏洞评分数据列表,φ_y表示指向节点对应的漏洞评分,且$\varphi_y\in\varphi$,ω_{xy}表示从源节点到指向节点边的权值. 在处理如图2中漏洞数据时,存在很多8、9分的高危漏洞主机,反映出整个内部网络系统十分脆弱,因此在由漏洞评分转为权值计算的时候,需要保持此种特性. 利用计算后的权值,作为奖励矩阵的值.

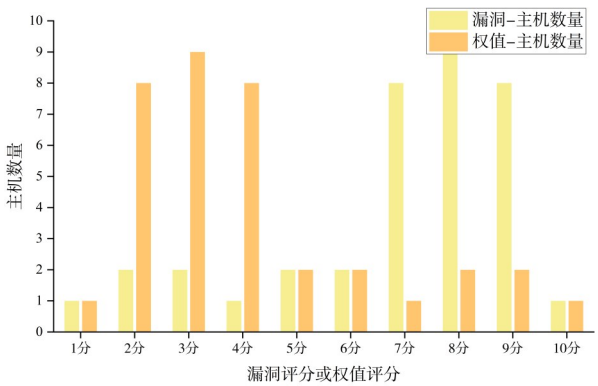


图2 漏洞映射权值数据

4.2.2 改进强化学习Q-learning算法实现

(1) 修改Q-learning算法中的$\max_{a'}Q(s',a')$为$\min_{a'}Q(s',a')$

由于对主机漏洞评分进行了权值转换,算法目标变为对期望最小的路径求最优解,因此实验中就需要将$\max_{a'}Q(s',a')$修改为$\min_{a'}Q(s',a')$,如式(3)所示. 假设当前真实奖励值$R+\gamma\min_{a'}Q(s',a')$小于当前期望值$Q(s,a)$,说明对于当前状态下的行动期望过高,通过状态价值函数减少期望,同理当期望过低时,应该存在如下关系$R+\gamma\min_{a'}Q(s',a')>Q(s,a)$,则状态价值函数期望增加. 综上所述,修改后的价值函数每次选择最小期望值,但是仍然保持修改后状态价值函数为收敛函数.

$$Q(s,a)\leftarrow Q(s,a)+\alpha\left(R+\gamma\min_{a'}Q(s',a')-Q(s,a)\right) \tag{3}$$

(2) 调整Q-learning算法中参数γ取值范围

本文实验采取在不考虑路径长短,只考虑路径代价的情况下,实现基于强化学习的类免疫动态攻击路径生成,所以应保持$\gamma=1$,改进后的Q-learning算法的状态价值函数如式(4)所示.

$$Q(s,a)\leftarrow Q(s,a)+\alpha\left(R+\min_{a'}Q(s',a')-Q(s,a)\right) \tag{4}$$

(3) 优化Q-learning算法中循环结束条件

在Q-learning算法中,循环结束条件是循环次数等于设定参数,然而对于有些简单网络,并不需要达到设定值,就可以找到最优攻击路径;同样的对于有些较为复杂的网络拓扑图,循环一定次数后并不能得到最优攻击路径. 为了解决以上两种问题,需要对结束条件进行优化.

首先在间隔一定次数的强化学习之后,对本次更新后的Q-table进行遍历,以寻找当前的最优攻击路径. 如果在某几次强化学习内随机挑选多次遍历路径,发现路径长度不变,那么可以认为该Q-learning算法面对实验环境已经具有了成熟的经验体系. 面对一个未知网络拓扑图,在并不知道有多少节点,多少条有向边的情况下,随意设定episodes值可能会影响路径探测结果,而通过本文提出的改进方案,可以加快验证最后结果是否趋于稳定.

(4) 改进后Q-learning算法

Q-learning算法改进后的详细步骤如算法1所示.

算法1 改进后Q-learning算法

输入:起始节点、目标节点、奖励矩阵、网络拓扑信息、学习率、ε-贪婪算法参数
输出:最优攻击路径
初始化Q-table
for e = 1 to episodes:
  初始化状态s;
  repeat:
  根据价值函数Q,通过策略(ε-贪婪算法)在状态s下选择动作a;
  执行动作a,获得奖励R以及下一状态s';
  $Q(s,a)\leftarrow Q(s,a)+\alpha\left(R+\min_{a'}Q(s',a')-Q(s,a)\right)$
    $s\leftarrow s'$
  until s 为终止条件
  if(指定范围遍历Q-table,节点长度趋于稳定):
    break for;
end for

ε-贪婪算法通过临时产生一个随机数字与贪婪参数比较,若随机数字大于或等于参数,则根据Q-table正常选择下一动作,若参数大于随机数字,则下一动作随机选择当前状态下的任意合理动作. ε-贪婪算法的存在,使得Q-table中动作在多方向上更新速率更快,而且存在一定几率的随机性.

4.2.3 解决最优攻击路径环路问题

最优攻击路径存在的环路问题主要有两种类型. 第一种是Q-table中期望值大部分为初始值,依据最小期望值遍历最优攻击路径时,通过随机方向选择动作,易出现未知环路,其原因是强化学习训练次数太少导致,随着网络拓扑图中节点数量增加、有向边增加,应该将强化学习次数增多,本实验中的结束条件遍历节点长度连续相同次数应相应增加. 另一种则是两个或两个以上经过强化学习训练后的节点,可能互有最小

期望值,其形成原因主要是环境反馈值引起,传统的下一动作选择方案依据分治算法,递归解决从 Q-table 中选择最小期望值的子问题,当两节点互有最小期望值时,在选择下一动作过程中,如果只考虑局部最优,忽略全局最优问题,就会形成死循环. 解决思路是在强化学习寻找最优路径时,如果本次选择的动作所指状态存在当前攻击路径中,则放弃本次最优动作,并且将该动作临时从当前状态可选动作集合中抛弃,然后继续寻找下一最优动作.

实验中还引入了违规动作集合,其作用是记录不可达攻击目标的路径,例如在网络拓扑图中某支路上不存在攻击目标,那么在强化学习的时候就记录下该支路,之后用于参照数据禁止强化学习算法往此方向学习.

4.3　攻击图构建

攻击图存在的目的是帮助显示获取的最优攻击路径,方便直观地看出内部网络脆弱性. 探寻内网最优攻击路径的主要难点之一就是建立合理而简洁的攻击图.

在常规攻击图中不考虑通信规则限制的情况下,网络通信是双向进行的,此时构建无权值、无向边攻击图是合理的. 但本文中需要以权值的方式利用网络主机信息,且由于网络通信方向表示了通信双方信息发送起始点,攻击者需要先后对相应节点进行利用,因此需要构建难度更大的含权值有向边攻击图.

5　实验与分析

5.1　实验需求

基于强化学习算法的最优攻击路径方案,需要预先搭建目标内网,模拟企业网络内部构建,进行方案验证. 内网中要求存在 DHCP(Dynamic Host Configuration Protocol)服务器、文件服务器、FTP(File Transfer Protocol)服务器、Web 服务器等必要服务,且需要各个主机终端开放必要端口和漏洞配置. 实验目的是在图 3 所示网络环境搭建的基础上,对本文方案各个功能模块进行逐一验证.

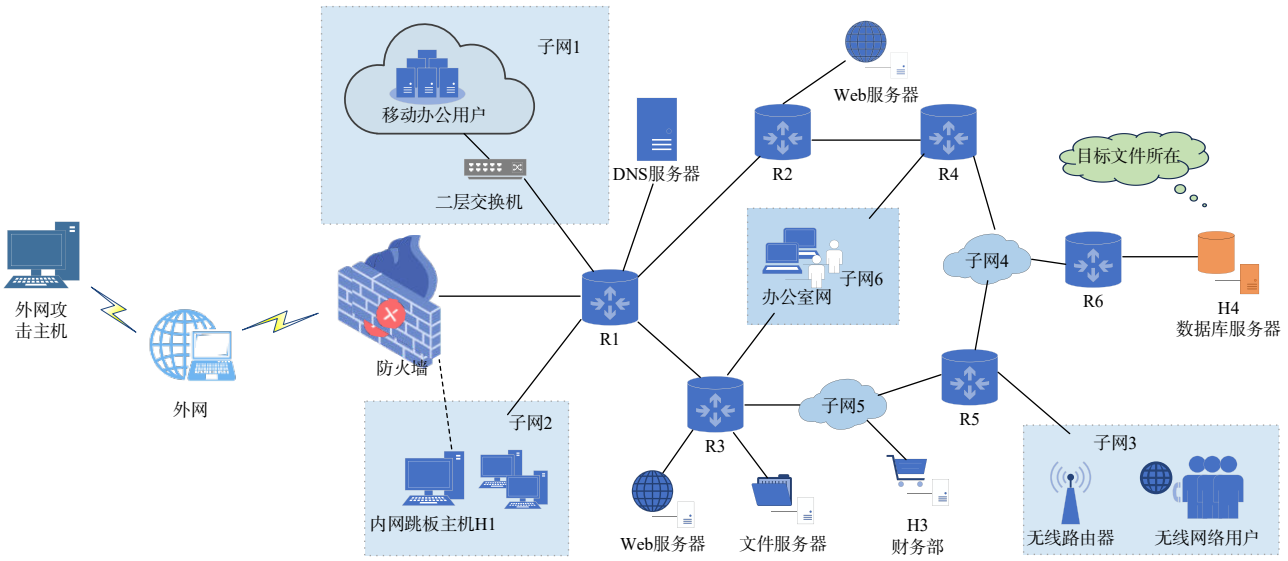


图 3　最优攻击路径方案实验网络拓扑

5.2　实验环境

与实际内网环境下的网络攻击相比,采用虚拟机方式模拟网络攻击可以更方便的进行网络配置,但存在一定的限制,虚拟机的性能受到主机资源限制,无法模拟大规模、复杂的攻击场景,同时,虚拟机环境中可能无法还原实际内网中存在的网络安全设备和防火墙,本文方案通过精心设计网络拓扑与攻击环境,尽可能保证虚拟环境与实际环境接近,保障实验的准确性.

模拟搭建企业内部网络时,通过在 Windows10 系统上采用虚拟机 VMware 软件实现. 外网攻击主机所使用的内部网络跳板主机 H1 为本机,目标主机 H4 为启用数据库服务器的 Ubuntu 主机,在虚拟机中通过六台 Windows2003 系统构建六个路由器(R1、R2、R3、R4、R5、R6)和各种类型服务,路由器之间通过 RIP(Routing Information Protocol)协议进行路由,并利用 H2 和 H3 充当企业内网中的办公主机.

实验中通过虚拟机中的主机设置添加虚拟网卡,对路由器终端网络接口进行扩展,可以实现多端口情况下的路由服务. 实验的内网环境中需要模拟很多服务类型,在本次实验中共实现四种服务:两个 Web 服务(连接 R3 的 Web 服务网络地址为 60.60.60.60:80,连接 R2 的 Web 服务网络地址为 192.1.1.2:80)、一个 DNS(Domain Name System)服务(连接 R1)、一个文件服务

器(连接R3,搭建在IIS(Internet Information Services)服务上的FTP服务器,服务器网络地址为60.60.60.60:21)和一个数据库服务器(位于目标主机H4上).

经过以上网络拓扑环境搭建、路由协议配置、服务器配置等步骤就可以基本实现对基于强化学习的最优攻击路径方案的验证.

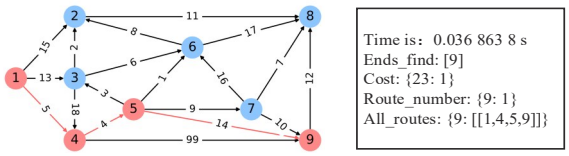
5.3 *Q*-learning算法准确性实验

首先对单条最优攻击路径的情况进行实验,人为构造网络拓扑图如4(*a*)所示,当从节点1到节点9,理论上只有一条最优攻击路径,路径经过的边的权值和为23.单条最优攻击路径实验结果如图4(*b*)所示,图中给出路径经过边的权值的总和为23,且最优攻击路径为一条.

在单条最优攻击路径实验的基础上将节点4到节点9的权值由原来的99改为18,如图5(*a*)所示,理论上最优攻击路径有1→4→5→9和1→4→9两条,多条最优攻击路径实验结果图如图5(*b*)所示,最优攻击路径为两条,且路径权值总和均为23.

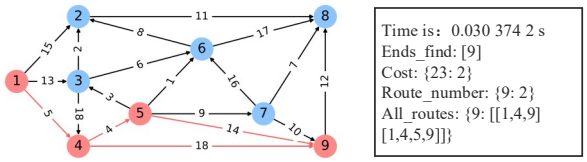
在进行存在环路攻击路径实验时,在单条最优攻击路径实验的基础上将节点1到节点4添加权值为1的环路,网络拓扑图如图6(*a*)所示,最优路径结果如图6(*b*)所示,最优攻击路径只有1→4→5→9且权值总和为19.

综上所述,*Q*-learning算法在面对单条、多条最优攻击路径以及存在环路的情况下均具有优秀的准确性.



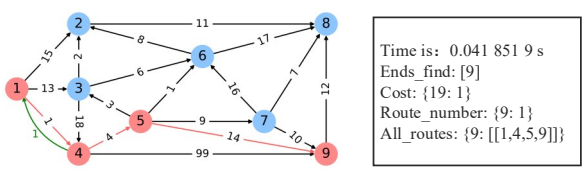
(*a*) 网络拓扑 (*b*) 扫描结果

图4 单条最优攻击路径实验



(*a*) 网络拓扑 (*b*) 扫描结果

图5 多条最优攻击路径实验



(*a*) 网络拓扑 (*b*) 扫描结果

图6 存在环路最优攻击路径实验

5.4 *Q*-learning算法效率实验

在本节中,测试在图7所示的条件下,本文方法每回合的时间消耗以及*Q*值收敛情况,并与文献[17]与文献[18]的方法进行对比.其中,对比方法的奖励矩阵构建均采用CVSS评分,本文方法采用4.2.1小节方法,同时文献[17]的学习率采用退火模型,文献[18]和本文学习率设置为0.2,其他参数保持一致.此条件下的最优攻击路径为H1→H4→H7→H10.

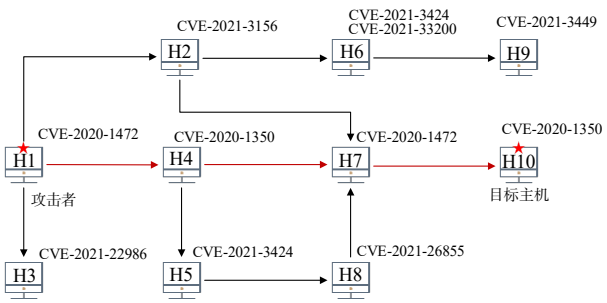
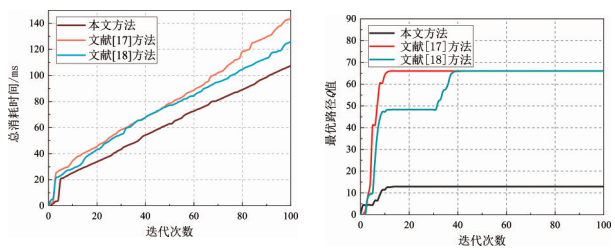


图7 效率对比网络拓扑

图8(*a*)描绘了三种方法在迭代次数100次时的总时间消耗,可以明显看出本文方法在效率上明显优于文献[17,18]所述方法,同时曲线更加平滑,说明每次迭代消耗时间较为稳定,原因是采用权值转换寻找最小*Q*值的方法可以减少对非关键路径的探索,减少时间消耗.图8(*b*)描绘了最优路径的*Q*值收敛情况,文献[17,18]所述方法采用最大*Q*值代表攻击路径,本文方法采用最小*Q*值代表攻击路径,相比文献[17,18]所述方法,可以快速收敛关键路径的*Q*值,减少计算量,同时避免如图所示因参数不当或迭代次数不够陷入局部最优的情况.本文方法在消耗时间与收敛稳定性方面优于文献[17,18]所述方法,因此本文方法更具有实用性.



(*a*) 消耗时间 (*b*) *Q*值收敛情况

图8 效率对比实验结果

对于大型网络拓扑条件下的运行效率,本文对网络拓扑图中节点从100个开始,每次递增100个节点,记录100个到2 600个节点进行实验的数据,统计效率实验中拓扑图节点数量、边的数量、强化学习算法寻找最优攻击路径总耗时,并将其转化为折线图如图9所示.

从折线图可以看出本文改进的*Q*-learning算法,可以

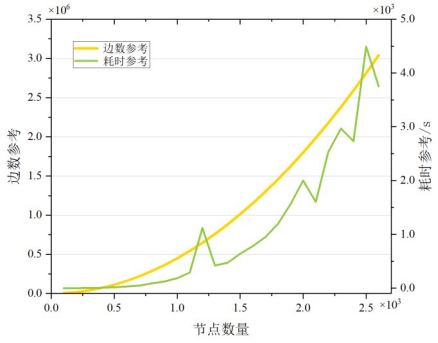


图9　Q-learning效率实验结果

在大型内网最优攻击路径问题上,极大地缩短路径寻找所用时间. 对于大型、复杂的网络拓扑图,广度优先搜索和深度优先搜索非常缓慢,并且随着最优攻击路径越来越深,广度优先搜索和深度优先搜索的时间呈指数增长. Dijkstra算法虽然消耗时间短,但只能给出1条最优攻击路径. 所以使用Q-learning算法进行最优攻击路径探索,将是在时间和结果上均优越于以上所提到的方法.

5.5　基于Q-learning算法的最优攻击路径方案实验

基于Q-learning的最优攻击路径方法验证实验步骤如下:

步骤1:按照5.2节实验环境描述搭建实验环境;

步骤2:在内网跳板主机H1上进行网络信息获取,读取路由表信息进行网段发现;

步骤3:利用Nmap的Ping命令进行存活主机扫描,根据扫描结果进行网络拓扑构建,如图10所示;

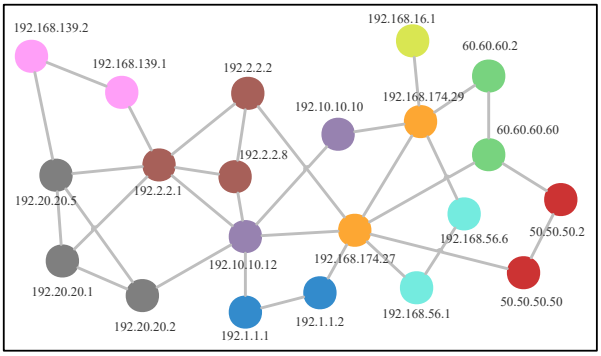


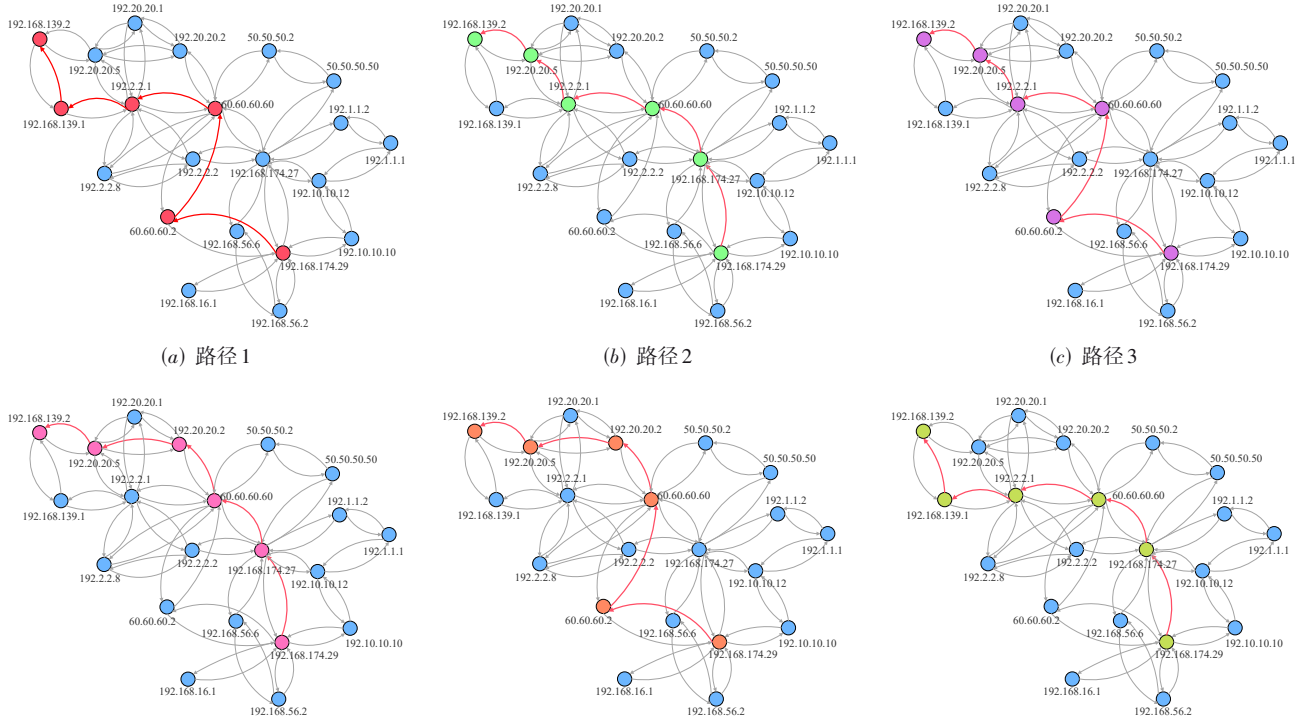
图10　网络拓扑图

步骤4:进行存活主机端口扫描和存活主机漏洞扫描,并对漏洞扫描结果进行漏洞量化和权值计算;

步骤5:运用Q-learning算法最优攻击路径寻找,根据结果构建动态攻击图.

Q-learning算法最优攻击路径实验结果如图11(a)~(f)所示. 从图中可以看出经过执行强化学习算法发现6条从H1(192.168.174.29)到H4(192.168.139.2)的最优攻击路径.

实验结果表明,6条攻击路径的危害是严重的,路径中重复率最高的60.60.60.60主机是内网中的脆弱



(a) 路径1　　(b) 路径2　　(c) 路径3

(d) 路径4　　(e) 路径5　　(f) 路径6

图11　构建攻击图结果

点,需要对其进行进一步的全面安全检测. 在网络安全维护过程中,需要对路径主机特殊端口进行选择性关闭,并升级相关安全插件,保证60.60.60.60的安全程度足够高,则6条攻击路径攻击代价都将会增大,内部网络安全等级也将全面提高.

6 结束语

为维护内部网络安全,本文介绍了一种基于强化学习的自免疫动态攻击生成方法. 本文方法首先对未知内部网络收集内部网络脆弱信息,并构建带权值的攻击图解决了内网脆弱信息利用率低的问题,并且通过对Q-learning算法的参数进行调整,优化了路径生成的效率. 通过实验表明:优化后的Q-learning算法可以适应网络的复杂情况,高效地发现内部网络中存在的脆弱点,在兼顾寻找最优攻击路径的准确性与高效性的同时,解决了内部网络中可能存在多条最优路径以及环路的问题,具有理论意义与实际价值. 未来,我们将结合博弈论,使得模型适用于复杂网络攻防场景[24~26],进一步提升内网自免疫安全防护能力.

参考文献

[1] 胡浩, 叶润国, 张红旗, 等. 基于攻击预测的网络安全态势量化方法[J]. 通信学报, 2017, 38(10): 122-134.
HU H, YE R G, ZHANG H Q, et al. Quantitative method for network security situation based on attack prediction [J]. Journal on Communications, 2017, 38(10): 122-134. (in Chinese)

[2] 闫峰, 刘淑芬, 冷煌. 基于转换的攻击图分析方法研究[J]. 电子学报, 2014, 42(12): 2477-2480.
YAN F, LIU S F, LENG H. Study on analysis of attack graphs based on conversion[J]. Acta Electronica Sinica, 2014, 42(12): 2477-2480. (in Chinese)

[3] 叶子维, 郭渊博, 王宸东, 等. 攻击图技术应用研究综述[J]. 通信学报, 2017, 38(11): 121-132.
YE Z W, GUO Y B, WANG C D, et al. Survey on application of attack graph technology[J]. Journal on Communications, 2017, 38(11): 121-132. (in Chinese)

[4] DACIER M, DESWARTE Y. Privilege Graph: An extension to the typed access matrix model[M]//Computer Security — ESORICS 94. Berlin: Springer, 1994: 319-334.

[5] WANG L Y, YAO C, SINGHAL A, et al. Interactive analysis of attack graphs using relational queries[M]//Data and Applications Security XX. Berlin: Springer, 2006: 119-132.

[6] RAMAKRISHNAN C R, SEKAR R. Model-based analysis of configuration vulnerabilities1[J]. Journal of Computer Security, 2002, 10(1/2): 189-209.

[7] RITCHEY R W, AMMANN P. Using model checking to analyze network vulnerabilities[C]//Proceeding 2000 IEEE Symposium on Security and Privacy. Piscataway: IEEE, 2002: 156-165.

[8] 臧艺超, 周天阳, 朱俊虎, 等. 领域独立智能规划技术及其面向自动化渗透测试的攻击路径发现研究进展[J]. 电子与信息学报, 2020, 42(9): 2095-2107.
ZANG Y C, ZHOU T Y, ZHU J H, et al. Domain-independent intelligent planning technology and its application to automated penetration testing oriented attack path discovery[J]. Journal of Electronics & Information Technology, 2020, 42(9): 2095-2107. (in Chinese)

[9] 李庆朋, 王布宏, 王晓东, 等. 基于最优攻击路径的网络安全增强策略研究[J]. 计算机科学, 2013, 40(4): 152-154.
LI Q P, WANG B H, WANG X D, et al. Approach on network security enhancement strategies based on optimal attack path[J]. Computer Science, 2013, 40(4): 152-154. (in Chinese)

[10] 杨本毅. 基于攻击图的渗透测试方法[J]. 电子科技, 2019, 32(10): 75-78.
YANG B Y. Research on corrosion detection technology of power system grounding grid[J]. Electronic Science and Technology, 2019, 32(10): 75-78. (in Chinese)

[11] NGUYEN T T, REDDI V J. Deep reinforcement learning for cyber security[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(8): 3779-3795.

[12] KAELBLING L P, LITTMAN M L, MOORE A W. Reinforcement learning: A survey[J]. Journal of Artificial Intelligence Research, 1996, 4: 237-285.

[13] CODY T, RAHMAN A, REDINO C, et al. Discovering exfiltration paths using reinforcement learning with attack graphs[C]//2022 IEEE Conference on Dependable and Secure Computing (DSC). Piscataway: IEEE, 2022: 1-8.

[14] 曾庆伟, 张国敏, 邢长友, 等. 基于分层强化学习的智能化攻击路径发现方法[J]. 计算机科学, 2023, 50(7): 308-316.
ZENG Q W, ZHANG G M, XING C Y, et al. Intelligent attack path discovery based on hierarchical reinforcement learning[J]. Computer Science, 2023, 50(7): 308-316. (in Chinese)

[15] CLIFTON J, LABER E. Q-learning: Theory and applications[J]. Annual Review of Statistics and Its Application, 2020, 7: 279-301.

[16] 李腾, 曹世杰, 尹思薇, 等. 应用Q学习决策的最优攻击路径生成方法[J]. 西安电子科技大学学报, 2021, 48(1): 160-167.

LI T, CAO S J, YIN S W, et al. Optimal method for the generation of the attack path based on the *Q*-learning decision[J]. Journal of Xidian University, 2021, 48(1): 160-167. (in Chinese)

[17] 胡昌振, 陈韵, 吕坤. 一种基于*Q*学习的最佳攻击路径规划方法: CN107317756A[P]. 2017-11-03.

[18] YOUSEFI M, MTETWA N, ZHANG Y, et al. A reinforcement learning approach for attack graph analysis[C]// 2018 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). Piscataway: IEEE, 2018: 212-217.

[19] JANG B, KIM M, HARERIMANA G, et al. *Q*-learning algorithms: A comprehensive classification and applications [J]. IEEE Access, 2019, 7: 133653-133667.

[20] 胡玮. 网络拓扑自动发现[D]. 成都: 电子科技大学, 2012.
HU W. Automatic Discovery of Network Topology[D]. Chengdu: University of Electronic Science and Technology of China, 2012. (in Chinese)

[21] 刘杰, 王清贤, 罗军勇. 一种基于ICMP的逻辑层网络拓扑发现与分析方法[J]. 计算机应用, 2008, 28(6): 1498-1500.
LIU J, WANG Q X, LUO J Y. ICMP-based method for logical network topology discovery and analysis[J]. Journal of Computer Applications, 2008, 28(6): 1498-1500. (in Chinese)

[22] PANSIOT J J, GRAD D. On routes and multicast trees in the Internet[J]. ACM SIGCOMM Computer Communication Review, 1998, 28(1): 41-50.

[23] DONNET B, FRIEDMAN T. Internet topology discovery: A survey[J]. IEEE Communications Surveys & Tutorials, 2007, 9(4): 56-69.

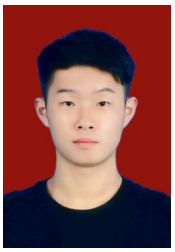
[24] 陈晋音, 胡书隆, 邢长友, 等. 面向智能渗透攻击的欺骗防御方法[J]. 通信学报, 2022, 43(10): 106-120.
CHEN J Y, HU S L, XING C Y, et al. Deception defense method against intelligent penetration attack[J]. Journal on Communications, 2022, 43(10): 106-120. (in Chinese)

[25] 张涛, 张文涛, 代凌, 等. 基于序贯博弈多智能体强化学习的综合模块化航空电子系统重构方法[J]. 电子学报, 2022, 50(4): 954-966.
ZHANG T, ZHANG W T, DAI L, et al. Integrated modular avionics system reconstruction method based on sequential game multi-agent reinforcement learning[J]. Acta Electronica Sinica, 2022, 50(4): 954-966. (in Chinese)

[26] WU Z N, TIAN L Q, WANG Y, et al. Network security defense decision-making method based on stochastic game and deep reinforcement learning[J]. Security and Communication Networks, 2021, 2021: 1-13.

作者简介

李　腾　男, 1991年1月出生于江苏省徐州市. 现为西安电子科技大学网络与信息安全学院副教授、博士生导师. 主要研究领域包括攻击检测与溯源、无人机系统和网络安全、漏洞挖掘与修复等. 中国电子学会会员编号:E190029987M.
E-mail: tengli@xidian.edu.cn



唐智亮　男, 2001年2月出生于安徽省铜陵市. 现为西安电子科技大学网络与信息安全学院硕士研究生, 研究方向为网络与信息安全.
E-mail: 1719468161@qq.com

马　卓(通讯作者)　男, 1980年12月出生于陕西省延安市, 现为西安电子科技大学网络与信息安全学院华山学者特聘教授、博士生导师, 主要从事人工智能、无人系统安全、隐私计算、无线网络安全等方面的研究. 中国电子学会会员编号:E190029397M.
E-mail: mazhuo@mail.xidian.edu.cn

马建峰　男, 1963年出生于陕西省西安市, 现为西安电子科技大学网络与信息安全学院教授, 博士生导师, 研究方向为网络安全、系统安全、数据安全和无人机安全. 中国电子学会会员编号:E190004733F.
E-mail: jfma@mail.xidian.edu.cn