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

臧艺超[①] 周天阳*[①] 朱俊虎[①②] 王清贤[①]

[①](数学工程与先进计算国家重点实验室(信息工程大学) 郑州 450001)

[②](国家数字交换系统工程技术研究中心 郑州 450001)

摘 要：攻击路径发现是自动化渗透测试领域的重要研究方向。该文综合论述了领域独立智能规划技术在面向自动化渗透测试的攻击路径发现上的研究进展及应用前景。首先介绍了攻击路径发现的基本概念并按照技术原理将其划分为基于领域相关和领域独立规划技术的攻击路径发现方法。然后介绍了领域独立智能规划算法，包括确定性规划算法、非确定性规划算法和博弈规划的技术原理和发展状况并就各类方法在攻击路径发现中的应用进行了综述。接着分析总结了渗透测试过程的特点，对比了领域独立智能规划算法应用在面向自动化渗透测试的攻击路径发现时的优缺点。最后对攻击路径发现将来的发展方向进行了总结和展望，希望对未来进一步的研究工作有一定的参考价值。

关键词：领域独立智能规划技术；自动化渗透测试；攻击路径发现

中图分类号：TN915.08; TP309 **文献标识码**：A **文章编号**：1009-5896(2020)09-2095-13

DOI: 10.11999/JEIT191056

Domain-Independent Intelligent Planning Technology and Its Application to Automated Penetration Testing Oriented Attack Path Discovery

ZHANG Yichao[①] ZHOU Tianyang[①] ZHU Junhu[①②] WANG Qingxian[①]

[①](*State Key Laboratory of Mathematical Engineering and Advanced Computing*, *Information & Engineering University*, *Zhengzhou* 450001, *China*)

[②](*National Engineering Technology Research Center of the National Digital Switching System*, *Zhengzhou* 450001, *China*)

Abstract: Attack path discovery is an important research direction in automated penetration testing area. This paper introduces the research progress of domain independent intelligent planning technology and its application to the field of automated penetration testing oriented attack paths discovery. Firstly, the basic concept of attack path discovery is introduced and the related algorithms are divided into domain-specific and domain-independent intelligent planning based attack path discovery algorithms separately. Secondly, the domain-independent planning algorithms are classified into deterministic planning, uncertain planning and game planning, where each of which is described from principle, development and application aspect in detail. Thirdly, this paper summarizes the characteristics of automated penetration testing and compares the advantages and disadvantages of domain independent intelligent planning algorithms adopted in automated penetration testing. Lastly, the development of automated penetration testing oriented attack path discovery is prospected. It is hoped that this paper could contribute future research works on attack path discovery.

Key words: Domain independent intelligent planning technology; Automated penetration testing; Attack path discovery

收稿日期：2019-12-31；改回日期：2020-03-17；网络出版：2020-07-21

*通信作者： 周天阳 aipteamzhouty@aliyun.com

基金项目：国家自然科学基金(61502528)

Foundation Item: The National Natural Science Foundation of China (61502528)

1 引言

网络信息技术在改善政府、企业等组织工作方式，提高工作效率的同时，使其面临的安全威胁与日俱增，尤其是近几年出现的高级可持续威胁等使得网络空间安全问题更加严峻。以入侵检测为代表的网络防护手段能够有效检测网络攻击行为，实现主机防护，但是这些防护措施是从防御者的角度出发被动地维护网络系统的安全稳定运行，并没有从攻击者的角度主动对网络系统的安全性进行检测。渗透测试通过发现目标网络中潜在攻击路径达到发现网络脆弱性的目的[1]。该过程需要渗透测试专家根据个人经验对目标网络进行信息获取、判断并调用攻击载荷来验证渗透测试方案的可行性，从而达到发现网络脆弱性的目的。但是渗透测试过程存在大量重复流程，耗费大量的时间和精力，加剧了渗透测试的代价，自动化渗透测试(automated penetration testing)为解决该问题提供了思路。

自动化渗透测试通过对目标网络、主机的自动化分析能够发现目标网络、主机潜在的脆弱性，并调用攻击载荷进行脆弱性验证[2,3]。目前成熟的自动化渗透测试工具包括APT2渗透测试套件、Autosploit渗透测试工具和Awesome-Hacking-Tools等。这些渗透测试工具提高了渗透测试效率，但仍存在一定问题，具体原因如下：从主机层面，这些渗透测试工具只是单纯地集成了已有的网络攻击工具，缺乏网络攻防知识推理能力，不能根据目标主机状态智能化选取攻击载荷并配置载荷参数进行渗透测试；从网络层面，这些渗透测试工具大多针对单个主机进行脆弱性评估，无法对整个目标网络的脆弱性进行评估，缺乏发现目标网络潜在攻击路径的能力。综上所述，实现攻防知识推理，智能发现攻击路径是实现自动化渗透测试的关键。典型的自动化渗透测试系统主要由渗透测试引擎和自动化渗透测试框架两部分组成。渗透测试引擎主要包含知识推理模块和智能规划等模块，负责根据网络状态发现攻击路径，实现智能决策，选择有效的攻击载荷并配置载荷参数；自动化渗透测试框架集合多种渗透测试工具，由渗透测试引擎驱动，负责与实际网络场景的数据交互和格式化反馈信息。本文着重研究自动化渗透测试系统中用于发现目标网络攻击路径的智能规划模块。

文章余下部分的结构如下：第2节介绍了面向自动化渗透测试的攻击路径发现的基本概念及其技术原理，包括基于领域相关智能规划技术的攻击路径发现和基于领域独立智能规划技术的攻击路径发现；第3节从确定性规划、非确定性规划和博弈规划3个角度介绍了现有领域独立智能规划技术的模型及相关研究进展；第4节分析了渗透测试过程的特点并从不同维度对比分析了现有领域独立智能规划算法在攻击路径发现时的优缺点；第5节分析了当前面向自动化渗透测试的攻击路径发现技术面临的问题与挑战，并指出潜在研究方向；最后总结全文。

2 基本概念及分类

本节首先介绍渗透测试和攻击路径发现的基本概念并就测试场景进行分类介绍，然后针对现阶段渗透测试存在的缺陷指出自动化渗透测试研究的必要性，并分类介绍面向自动化渗透测试的攻击路径发现技术。

2.1 渗透测试及攻击路径发现的基本概念

目前学术界对渗透测试没有明确统一的定义，McDermott将渗透测试过程阶段化描述，划分为明确渗透测试目标、信息收集、假设目标缺陷、确认目标缺陷、扩展目标缺陷、消除目标缺陷6个阶段[4]。虽然学术界对渗透测试没有明确定义，但业界普遍认为渗透测试是一种通过模拟恶意攻击者的技术和方法，挫败目标系统安全控制措施，取得访问控制权限，并发现业务安全隐患的一种安全测试与评估方式。按照实际应用场景将渗透测试划分为黑盒测试、白盒测试和灰盒测试[5]。黑盒测试是指在没有任何目标网络内部拓扑等相关信息的条件下，发现目标网络中存在的一些已知或者未知安全漏洞的过程；白盒测试是指在完全了解目标网络环境中主机和网络拓扑信息条件下发现目标网络或主机中存在安全漏洞和脆弱性的过程；灰盒测试是指在部分了解目标系统主机和网络拓扑信息条件下，通过信息收集、漏洞利用等动作来评估目标网络系统安全性的过程。攻击路径是目标网络中存在的可以被攻击者用于获取特定资产的漏洞序列，而攻击路径发现，作为自动化渗透测试的核心，是从目标网络中发现所有这些漏洞序列的技术[6]。自动化渗透测试旨在提高渗透测试的自动化程度，通过智能分析目标网络环境，发现攻击路径，调度渗透测试工具以实现目标网络的脆弱性检验。

2.2 攻击路径发现技术分类

面向自动化渗透测试的攻击路径发现是人工智能技术在网络空间安全领域中的重要应用，现有渗透测试方法主要依赖于专家经验，不能智能化分析网络态势和发现网络系统潜在攻击路径，利用智能规划技术提高攻击路径发现的自动化程度是实现自动化渗透测试的关键。传统意义上的智能规划技术是在给定初始信息条件下，通过选择有效的动作序

列从而达到既定的目标状态，如机器人路径规划导航问题等[7]。从领域相关性角度，智能规划技术可以分为两大类：领域相关(domain-specific)智能规划技术和领域独立(domain-independent)智能规划技术[8]。领域相关智能规划是指只适用于特定领域的规划模型/算法，例如基于攻击图的攻击路径发现只适用于网络安全领域，主要分为两大类：基于状态攻击图的攻击路径发现[9–11]和基于属性攻击图的攻击路径发现[12,13]。状态攻击图的节点一般表示主机名称以及获取的用户权限等信息，边代表具体的原子网络攻击，其执行会引起全局目标系统状态的变迁。基于状态攻击图进行攻击路径发现时由于状态节点代表系统全局状态，而目标网络通常是极度复杂的，因而导致状态攻击图节点数目呈指数增长，不适用于大规模目标网络环境下的攻击路径发现。属性攻击图的节点表示原子攻击节点的前提或者结果。只有当所有与攻击节点相连的属性节点均满足时，原子攻击才能实施。属性攻击图在一定程度上简化了攻击图规模但较为抽象，难以应用于实际渗透测试的攻击路径发现。综上可知，攻击图技术虽然能够协助发现攻击路径，但应用在渗透测试领域仍存在以下问题：一是攻击图构建过程复杂，耗时长难以满足渗透测试的时间要求；二是构建的攻击图复用难，目标网络变化即需重新构建，造成计算资源浪费；三是攻击图的构建需要提前扫描获取目标网络完整信息，难以在灰盒和黑盒渗透测试中得到满足，从而导致基于攻击图的路径发现方法难以应用于面向自动化渗透测试的攻击路径发现。攻击图技术将目标网络信息与攻防知识信息紧耦合，降低了算法的可扩展性，因此需要领域独立智能规划技术进行知识表达与快速攻击路径发现。领域独立智能规划是一种通用技术，可以应用于多个领域，例如规划图、偏序规划、分层任务网络等确定性规划算法，概率模型、马尔可夫模型、部分观测的马尔可夫模型等非确定性规划算法以及静态博弈、演化博弈、Stackelberg博弈等博弈规划模型。面向自动化渗透测试的攻击路径发现是在已有网络知识结构下，对已知或者部分已知的网络系统进行脆弱性发现，然后通过信息采集、漏洞利用等手段实现目标网络自动化安全验证的过程。国内研究大多以攻击图等相关技术为基础的领域相关攻击路径发现算法为主，而基于领域独立智能规划技术的攻击路径发现研究较少，主要原因：一是领域独立智能规划作为一种通用技术主要被应用于飞行轨迹规划等任务性场景领域[14,15]，而攻击路径发现为网络安全领域的重要问题，虽然已有部分领域独立智能规划算法已经用于攻击路径发现[16,17]，但研究领域差异导致很多领域独立智能规划算法未被应用于攻击路径发现；二是领域独立智能规划算法应用于攻击路径发现需要将领域知识转化为领域独立智能规划算法的表达方式，增加了算法应用的难度。例如采用图规划、偏序规划等规划算法时需要将攻击路径发现的领域知识转化为规划领域定义语言PDDL (Planning Domain Definition Language)的表达形式，这种知识表示的转化增加了领域独立智能规划算法应用于攻击路径发现的难度。已有研究工作表明，领域独立智能规划算法与攻击路径发现相结合取得了显著的效果，因此有必要对现有领域独立智能规划技术及其在攻击路径发现中的应用前景进行总结，对比不同规划算法的原理、适用性及其优缺点，为面向自动化渗透测试的攻击路径发现问题提供解决思路。

3　领域独立智能规划技术及其在攻击路径发现中的应用

3.1　确定性规划技术及其在攻击路径发现中的应用

确定性规划算法是一类基于状态转移系统的领域独立智能规划方法，主要解决确定、静态、完全观测条件下的路径规划问题。基于确定性规划算法的攻击路径发现主要包含两个步骤：首先根据漏洞信息和网络场景信息将攻击路径发现转化为PDDL表示形式；然后利用确定性规划算法发现攻击路径。根据规划模型的不同将算法划分为基于规划图技术、基于偏序规划、基于分层任务网络的路径规划。

3.1.1　规划图模型及相关研究进展

基于规划图技术的路径发现通过构建“规划图”实现对路径的表示和搜索，该模型主要包含两类节点：一类是proposition节点用于描述当前可行状态；另一类是action节点用于描述当前可行动作[18]。路径规划过程由图扩展和解提取阶段交替进行。图扩展阶段采用前向搜索的方式从当前可行状态集合出发查找当前可用动作集合并进行扩展生成下一层的可行状态集合，然后判断生成的状态集合是否满足目标状态要求，并更新互斥proposition节点对和action节点对集合。为了能够快速搜索可行路径，需要判断同一层次内互斥的action或proposition节点对。解提取阶段依据互斥条件，首先对最后一层proposition节点集合进行判断，观察其是否包含目标状态集合。若包含，根据目标状态的前置条件更新目标状态集合并后向搜索直到获取可行路径，否则对当前规划图再次进行图扩展操作，进入下一轮迭代。在进行解提取的过程中需要根据互斥proposition节点对集合和互斥action节点对集合进行回溯约

束，获取非互斥路径。基于规划图模型的路径规划能够有效发现潜在路径，但存在求解效率不高的问题。后续学者针对该问题进行了深入研究。文献[19]将基于规划图的规划问题转化为布尔可满足性SAT问题并利用通用SAT求解器进行求解。Kambhampati等人[20]基于规划图后向搜索与CSP(Constraint Satisfaction Problem)搜索的相似性将规划图模型转化为约束可满足CSP问题进行规划求解，并指出在运行时间和内存消耗方面，CSP相对于SAT求解的测试效果更好。Baioletti等人[21]将满足性规划技术应用到规划图中达到缩减搜索空间的目的，从而提高搜索效率。

3.1.2 偏序规划模型及相关研究进展

基于偏序规划的路径发现通过维护动作之间的偏序关系和冲突关系使路径搜索更加有效[22,23]。不同于规划图，偏序规划仅约束了必须存在严格先后关系的动作，而对于其他动作之间的先后关系不做严格要求，任何满足偏序关系的规划路径都是可行路径。基于偏序规划的路径规划算法首先构建一个仅包含初始-终止状态节点的因果链，然后以后向搜索算法为基础从终止状态集合中选择子状态并查找能够满足该子状态的precondition动作，形成因果链并将动作添加到目标状态集合当中；然后遍历动作空间查找与因果链相冲突的动作，利用降级/升级(demoting/promoting)动作构成顺序约束，顺序约束表示约束的两个动作必须严格按照此顺序，否则不满足最终目标状态。不断迭代上述过程直到目标状态满足从而可以搜索得到所有可行路径。偏序规划过程由于需要多次遍历动作集合，因此算法复杂度较高。Younes和Simmons[24]在偏序规划的基础上，利用缺陷选择策略与启发式可达性分析相结合的方式提高路径规划发现率，形成了偏序因果链规划器VHPOP。文献[25]将偏序规划算法整合到前向搜索框架中，实现了最少承诺和全承诺之间的平衡，有效地解决了时序数值规划问题。针对多代理场景导致的行为交叉影响问题，Boutilier和Brafman[26]在偏序规划算法的基础上引入并发行为选择机制有效解决了并发行为之间的相互影响问题。

3.1.3 分层任务网络模型及相关研究进展

基于分层任务网络(Hierarchical Task Network, HTN)[27–29]的路径发现不同于其他规划算法，HTN基于任务(Tasks)和任务分解的方法实现路径发现。HTN方法中主要包含3类任务：原始任务(goal task)、非原子任务(compound task)、原子任务(primitive task)，其中原始任务表示渗透测试过程的目标状态，非原子任务表示任务分解过程中得到的不可直接执行的复合任务状态，原子任务表示任务分解过程中得到的可以直接执行的任务状态。基于分层任务网络的路径发现由3部分构成：原始任务描述、任务分解方法和约束条件。算法思想是对原始任务进行分解，获取得到子任务，如果子任务是原子任务，则该子任务可直接完成，否则继续分解，直到子任务能够完成从而推断出可行路径。该类算法的优点在于能够从高层抽象的角度对规划问题进行分析，同时可以忽略底层具体实现细节，易于理解，缺点是对待具体问题往往需要定制具体的问题分解方法及实现方法，自动化程度不高。Mu和Li[30]将分层任务网络规划算法应用到入侵检测中并提出了一种入侵响应决策算法，该算法主要由两部分组成：响应评估决策过程和响应时间决策过程。根据响应时间设置的不同，算法能够优化规划方案实现规划效率和响应代价的平衡。文献[31]将分层任务网络算法与树搜索算法相结合提出了一种对抗分层任务网络算法，解决了状态空间大、组合爆炸导致的路径发现效率低问题。文献[32]针对非确定性条件下大规模规划方案求解难的问题提出了一种规划算法YoYo，该算法将分层任务网络中的搜索控制策略与基于符号模型检测的规划技术相结合实现了非确定性情况下的大规模规划问题快速求解，实验结果表明该算法在求解效率上有较大提升。针对算法中存在的知识转化难且耗时的问题，文献[33]提出了一种知识学习算法HTNLearn，该算法以注释过的规划方案和任务集合作为输入，将任务分解方法的约束问题转化为CSP问题并采用加权的MAX-SAT求解器进行求解从而获取任务分解方法和动作行为空间。

本节对于基于确定性规划算法的路径发现技术从规划图、偏序规划、分层任务网络3个方面归纳了主流的确定性规划技术。这3类技术本质是利用搜索技术来遍历路径空间发现可行解。基于规划图、偏序规划都是利用后向搜索算法以目标状态为出发点进行搜索，直到初始条件满足要求为止，而分层任务网络规划算法是利用前向搜索算法以初始状态为出发点进行搜索，直到达到目标状态。区别于穷举搜索算法，这些规划算法在进行状态空间搜索过程中分别采用了不同的剪枝技术提高搜索效率。例如规划图利用了动作/状态对的互斥关系，偏序规划利用了动作之间的冲突关系等来提高搜索效率。确定性规划技术研究理论殷实、工具成熟，得到了广泛的应用。在自动化渗透测试领域，该类算法进行规划时需要获取目标网络所有信息，更适用于白盒渗透测试条件下的攻击路径发现。

3.2 非确定性规划技术及其在攻击路径发现中的应用

确定性规划算法主要解决静态、确定、完全可观测条件下的路径规划问题，但渗透测试问题往往是动态的、非确定性、部分观测条件下的路径发现，这导致经典规划算法失效，因此研究非确定性规划技术对实现攻击路径发现具有重要意义。领域独立的非确定性规划算法主要包含两个研究方向：一是在确定性规划算法研究的基础上通过引入概率不确定性实现非确定性条件下的路径发现; 二是通过构建马尔可夫规划模型实现对渗透测试过程不确定性的描述，从而实现非确定性条件下的路径发现。

3.2.1 概率条件下的确定性规划模型及相关研究进展

该类方法通过在确定性规划模型中引入概率分布刻画不确定性构建概率规划模型。基于概率规划模型的策略分析方法包含两大类：基于Determinizing技术的路径规划[34]和基于概率优化的路径规划[35]。

基于Determinizing技术实现非确定性条件下的路径规划包含两个步骤：首先是通过约束放松的方式将非确定性问题转化为多个确定性规划问题；然后利用确定性规划算法对转化后的规划问题进行求解。该方法能够有效地利用现有确定性规划研究成果实现非确定性条件下的路径发现。Cimatti等人通过符号模型检测的方式对非确定性规划问题进行求解，求解结果分为3种情况：弱规划解(weak plan)、强规划解(strong plan)和强循环规划解(strong cyclic plan)。其中弱规划解表示可能到达目标状态；强规划解表示一定能够到达目标状态；强循环规划解表示循环以一定概率停止，当停止时该解一定能够到达目标状态。Muise等人[36]针对完全可观测非确定性规划FOND(Fully Observable Non Deterministic)问题提出了基于状态关联的非确定规划器PRP(Planner for Relevant Policy)，该规划器依据状态关联信息构建强循环规划实现非确定条件下的路径规划。Muise等人[37]后续又对PRP进行了扩展，通过引入条件效果增强了PRP规划器的表达能力。李洋等人[38]针对强循环规划问题提出了基于最小期望权重的求解方法，该方法的主要思想是利用深度优先搜索算法求出规划问题的所有强循环规划解，再将强循环规划解分别转换成以状态到目标状态的期望权值为变元的线性方程组，最后使用高斯消元法求解方程组，从而找到最小期望权值强循环规划解。唐杰等人[39]考虑动作执行可逆的情况构建了新的不确定性规划模型然后通过构建规划子图和规划子树实现非确定性的路径规划。

基于概率优化方法的路径发现和基于攻击图技术的路径发现均是通过最大化目标函数实现路径发现。二者的区别在于作用的模型和优化目标不同。Kushmerick等人[40]通过引入概率分布来描述世界状态的非确定性，并据此实现了BURIDAN规划器最大化目标状态概率。Yoon等人[41]基于FastForward规划器，通过对概率规划问题的分解实现了一种动态重规划算法FF-Replan，该算法首先从多种可能的行为结果中随机选择一个确定的行为结果，并据此将非确定性规划问题转化为确定性规划问题，然后利用FF(FastForward)规划算法实现规划求解，并按照求解得到规划结果进行执行，如果遇到非预期状态则以该状态为初始状态迭代进行上述流程。Yoon等人[42]通过将非确定性规划问题分解为多个确定性规划问题，其中分解得到的确定性规划问题为可能行为结果的组合，然后利用确定性规划算法分别求解得到可行规划序列集合，再从该集合中对规划序列进行分析评估，得到最终规划序列。但是该方法由于存在组合爆炸问题导致求解复杂度过高，Issakkimuthu等人[43]在此基础上通过引入概率有益动作避免了大量无效动作的组合从而大大提高了算法那求解的效率。Bryce等人[44]在规划图技术的基础上通过构建启发函数实现非确定性规划，并将其归约为概率分布估计问题，并采用序贯蒙特卡洛算法进行求解。Trevizan等人[45]提出了一种基于占用测度的启发式搜索算法，该算法通过计算每一个动作的期望执行次数来最小化行为代价，最终实现非确定性规划。

3.2.2 马尔可夫模型及相关研究进展

基于马尔可夫模型实现非确定性规划主要包含两种方式：基于马尔可夫决策过程(Markov Decision Process, MDP)和基于部分观测的马尔可夫决策过程(Partially Observation Markov Decision Process, POMDP)。其中POMDP过程在MDP过程的基础上增加了状态观测的不确定性，但其策略分析算法却存在较大差异，因而规划决策效果也不同。

马尔可夫决策过程采用5元组表达形式$< S, A, P, r, \gamma >$。其中S为主机状态空间，A为攻击动作空间，$P(s,a,s') = Pr(s'|s,a)$为状态转移函数，$r(s,a)$为奖惩函数，γ为折扣因子。马尔可夫决策过程的目标是选取最优策略π最大化长期累积奖励$J^{\pi}(\mathrm{s}_0)$

$$\pi^* = \underset{\pi}{\operatorname{argmax}}\, J^{\pi}(\mathrm{s}_0), J^{\pi}(\mathrm{s}_0) = \mathrm{E}\left[\sum_{t=0}^{T-1} \gamma^t r_t | \mathrm{s}_0, \pi\right]$$

文献[46]将渗透测试过程形式化为MDP过程，其中动作空间由具体的漏洞组成，状态空间由攻击动作

及结果组成，奖励函数由常状态转移量与损失值组成，整个模型的目标是最小化期望损失值。针对马尔可夫决策过程，目前常用的求解策略包括动态规划、策略迭代、值迭代、蒙特卡洛以及TD(λ)等求解方式。基于策略迭代的求解方法主要包含策略评估(policy evaluation)和策略更新(policy improvement)两个步骤，在Banach不动点定理的保证下，该算法能够保证收敛到最优策略[47]。基于值迭代、TD(λ)和Q-learning的求解方式的核心思想都是基于差值引导的策略优化方案，区分点在于差值的计算方法不同。值迭代的策略更新基于最优Q值响应；TD(λ)的策略更新基于Q值的更新估值和当前估值的差值；Q-learning的策略更新基于最优更新估值与当前估值的差值。相对于通过值函数间接求解最优策略，策略梯度方法直接对目标收益函数求导获取最优梯度方向并利用梯度下降算法获取最优策略参数从而达到求解最优策略的目的[48]。为了提高求解效率，TRPO算法以KL散度为约束条件引入信赖域优化技巧提高目标函数求解效率[49]。为了解决策略梯度方法中的训练不稳定问题，Actor Critic方法[50]被提出，该方法通过引入优势函数(advantage function)达到提高训练稳定性的目的。文献[51]在Actor Critic算法的基础上通过引入异步多线程优化技巧提出了A3C算法，该算法让多个线程agent同时学习最优策略并进行交互达到快速优化收敛的目的。上述算法适用于状态空间小的情形，当状态空间较大时求解效率较低，难以满足实际需求。为了解决主机信息组合造成的状态空间爆炸问题，不少学者提出了大规模状态空间策略求解算法。在状态空间比较大的情况下，传统的基于状态-动作表的策略描述方法往往导致内存空间消耗过大且稀疏，最终导致策略收敛时间长。文献[52]提出了一种基于函数逼近的策略表示方法，使用包含参数向量$\boldsymbol{w}_k$的函数表示$Q(s,a)$值函数，在整个策略的计算过程不需要存储$Q(s,a)$值函数，只需要存储参数向量$\boldsymbol{w}_k$使得求解效率也大大提升。Taylor等人[53]利用LSSVR模型将含参值函数逼近问题转化为高维特征空间中的线性回归问题，在保证泛化能力的前提下有效地提高了算法地收敛速度。

部分马尔可夫决策过程采用7元组表达形式$<S,A,\mathcal{O},P,T,r,\gamma>$，增加了主机观测空间$\mathcal{O}$和状态观测函数$T(s,a,o)=Pr(o|s,a)$。文献[54]将渗透测试形式化为POMDP过程，其中主机状态空间由主机、应用程序、端口组合构成；动作空间从Metasploit渗透测试框架漏洞库中提取，主要包含操作系统扫描、端口扫描、漏洞利用3类动作；观测空间包含操作系统类型、获取主机权限等。为了实现路径规划，获取最优策略，需要对POMDP模型进行求解，目前对POMDP算法的研究包括精确算法和近似算法，精确算法理论上可以获得最优解，但由于计算复杂性随着问题规模呈指数增长，一般只适用于求解一些小规模问题，因此出现了许多求解POMDP的近似算法。Sondik在其博士论文中首次提出了POMDP模型的精确求解算法One-Pass算法，该算法利用了最优值函数是信念状态空间上的分段线性函数这一特性将信念状态空间进行划分然后分别求解每一划分区域上的最优行动从而获取最优策略[55]。Cheng等人[56]提出了一种线性支撑算法，采用松弛边界条件，将原问题转化为凸优化问题进行求解从而获取最优策略。由于POMDP模型求解效率低，Pineau等人[57]提出了PBVI算法实现POMDP模型近似求解，首先选取典型信念点集合，然后利用值更新策略更新这些信念点，最后对这些信念点进行最优策略选取，大大提升了求解效率。Liu等人[58]提出了一种信念点下界函数，可以有效提高算法收敛效率，进一步基于该下界函数进行剪枝提高算法求解效率。Kurniawati等人[59]利用边界限定技巧限制非最优区域的采样，提高采样效率并提出了SARSOP算法实现最优策略求解。上述求解方法只能进行单机决策规划，为了能够实现网络层面的攻击规划，文献[60]提出了4AL算法实现网络规划，主要包含3个阶段：首先分解目标网络为有向无环图；然后将图中全连通节点进行分解，将属于同一个子网的节点划分到一起，最后对每一个节点利用POMDP模型进行攻击规划，最后整合形成攻击路径。

由上述分析可知，Determinizing规划技术能够对路径规划中面临的非确定性问题进行分解，然后利用确定性规划算法进行求解。概率规划利用概率分布对渗透测试过程中非确定性进行描述，以最大化渗透成功概率为目标进行路径优化，而马尔可夫决策模型以最大化长期收益为目标，将渗透测试中的不确定性以期望收益的形式融入路径规划。非确定性规划技术能够有效刻画渗透测试过程的不确定性，使得路径规划的现实依据更强、成功率更高，因此具有良好的研究前景，但因算法求解复杂度高，难以扩展到大规模网络场景，限制了其实际应用，提高求解效率是该类方法未来研究的重点。

3.3 博弈模型及其在攻击路径发现中的应用

无论是确定条件还是非确定条件下的攻击路径发现都是从攻击者角度单方面进行的路径发现，没

有考虑防御方采取的防御策略及其对攻击路径发现的影响，在实际的渗透测试过程中往往伴随着攻防双方的博弈，因此研究攻防博弈条件下的路径规划对提高渗透测试的自动化程度具有重要意义[61]。基于博弈模型的攻击路径发现分为基于静态博弈模型和基于动态博弈模型的攻击路径发现两类，其中静态博弈模型假定攻防双方的策略选取及收益保持不变，而动态博弈模型摒弃了该假设，认为攻防双方的策略选择及收益可以是动态变化的，下面对基于上述两类博弈模型的攻击路径发现研究现状进行详细介绍。

3.3.1 静态博弈模型及相关研究进展

基于静态博弈模型的攻击路径发现以经典博弈模型为基础，结合渗透测试领域知识定义攻防双方的策略空间、收益损失函数。应用于攻击路径发现，对于防御方而言是最优防御策略，对于攻击方而言为最优路径攻击策略，该策略描述了针对当前攻防博弈状态及网络态势条件下的目标及动作选择，指导攻击路径发现。Lye等人[62]首次提出利用博弈理论分析计算机网络安全，将网络攻防交互过程视为二人随机博弈模型进行安全性分析从而达到发现网络潜在攻击路径和增强网络安全性的目的。姜伟等人[63]为了发现网络系统潜在攻击路径并制定有效的主动防御措施首先提出了网络防御图的概念，对攻防对抗策略的成本收益进行分析，然后在防御图模型的基础上结合成本效益分析构建攻防博弈模型，最后提出基于该模型的最优主动防御策略选取算法，帮助防御者发现潜在攻击路径并采取最优防御策略进行主动防御。王晋东等人[64]建立了非完全信息条件下的静态贝叶斯博弈主动防御模型，该模型详细考虑了攻击者和防御者类型并全面分析了攻防对抗条件下的混合策略贝叶斯均衡，将攻击者的混合均衡策略概率作为可能攻击路径，提出了基于防御效能的最优策略选取算法。王元卓等人[65]提出了随机博弈模型的网络攻防实验整体架构, 提出了基于网络连接关系、脆弱性信息的网络攻防博弈模型快速建模方法，并利用攻防模型可以对目标网络的攻击成功率、平均攻击时间、脆弱节点以及潜在攻击路径等方面进行安全分析与评价。Cui等人[66]提出了一种基于马尔可夫博弈理论的网络信息系统风险评估模型，考虑当前系统风险因素和将来可能出现的系统风险因素，对网络信息系统的安全性综合评估，从而发现系统潜在攻击路径。

3.3.2 动态博弈模型及相关研究进展

基于静态博弈模型的攻击路径发现假设攻防双方的攻防策略等稳定不变并以此为基础发现攻击路径，但是攻防对抗过程是一个动态的过程，防御策略及攻击效益等因素是动态变化的，基于静态博弈模型的攻击路径发现难以满足动态变化条件下的路径发现，因此不少学者开始研究基于动态博弈模型的攻击路径发现。Li等人[67]针对网络攻防技术的时间连续性问题通过将离散多阶段攻防博弈模型连续化实现非完全信息条件下的攻击路径发现达到控制网络信息系统风险的目的。黄健明等人[68]针对现有攻防博弈模型中的完全理性假设，从对抗过程中的有限理性条件出发，构建攻防演化博弈模型并提出了一种演化稳定均衡求解方法，发现网络系统潜在攻击路径并设计了最优防御策略选取算法。张恒巍等人[69]针对具有不完全信息约束的多阶段动态攻防过程，构建了多阶段攻防信号博弈模型，在此基础上设计了多阶段攻防博弈均衡求解算法，发现网络潜在攻击路径并给出了最优主动防御策略选取方法。朱建明等人[70]基于系统动力学提出了在信息不对称情况下的攻防演化博弈模型，该模型结合攻防效用函数实现对非合作演化博弈攻防过程的形式化描述能够有效发现网络潜在攻击路径。此外，不少学者开始研究利用Stackelberg模型描述渗透测试过程中非对称信息条件下的攻防对抗过程。文献[71]首次将Stackelberg博弈应用于面向web的移动目标防御，将web配置错误导致的漏洞利用问题转化为Stackelberg博弈问题，其中防御方为leader，攻击者为follower，通过权衡合法用户的正常访问与攻击者的恶意攻击引起的收益/损失，达到发现系统潜在脆弱路径和最优服务配置的目的。Yuan等人[72]为实现智能攻击环境下的工控系统最优弹性控制将攻防对抗过程形式化为一种多阶段的分层博弈模型，每一层均采用完全信息条件下的Stackelberg博弈实现最优控制从而达到发现工控系统的潜在脆弱路径，提高工控系统的弹性控制。

基于博弈模型的攻击路径发现主要从静态博弈和动态博弈两个角度介绍了博弈模型在攻击路径发现中的应用。基于静态博弈模型的攻击路径能够充分考虑防御者的防御措施发现攻击路径，提高了渗透测试的成功率。但需要攻防双方的攻击/防御策略稳定，因此更适用于网络环境已知，攻防手段固定条件下的渗透测试。动态博弈技术相比于静态博弈技术能够考虑攻防双方策略的动态变化，识别网络中存在的脆弱点和潜在的安全威胁，从而达到在动态对抗条件下发现攻击路径的目的。但是该类方法模型复杂且耗时严重导致应用受限，提高模型适用性，降低计算复杂度成为动态博弈模型的研究重点。

4 领域独立智能规划技术用于攻击路径发现的分析对比

领域独立智能规划方法应用于自动化渗透测试领域目前没有统一的归纳总结，各方法既有相关性，又存在差异性，因此有必要结合渗透测试自身的特点对算法进行综合对比分析。本节首先分析渗透测试过程的特点，然后以此为依据对比不同领域独立智能规划算法表现并分析其存在的优缺点。

4.1 渗透测试的特点

状态空间完备性：渗透测试的状态空间完备性是指渗透测试之前针对目标网络的掌握情况，根据掌握信息的多少，渗透测试分为3种：第1种是白盒渗透测试，指目标网络及主机配置信息、漏洞信息完全已知，需要根据完全信息规划出攻击路径实现渗透目的；第2种是灰盒渗透测试，指目标网络及主机配置信息、漏洞信息部分已知，需要根据已知的部分信息在探测网络信息的同时实现攻击路径发现；第3种是黑盒渗透测试，指目标网络及主机配置信息、漏洞信息完全未知，需要逐步探测目标网络信息，并根据部分获取的网络及主机配置信息进行规划，不断迭代直至达到渗透测试目的。在实际网络渗透测试过程中，根据观测角度不同，渗透测试又分为两类。从攻击者角度，由于代理及防火墙等安防设备的存在导致难以完全掌握目标网络信息，形成部分信息条件下的攻击路径发现；而从防御者角度，由于网络拓扑，主机配置等信息完全已知，形成完全信息条件下的攻击路径发现。

行为不确定性：实际的渗透测试中存在着大量的不确定性，某些攻击行为产生非预期的效果，例如：由于攻击载荷目标适用性或自身代码质量缺陷导致的攻击失败。这种不确定性广泛地存在于渗透测试过程中，如果忽略攻击行为的不确定性，攻击规划的复杂度将会大大降低，在一定程度上也可以保证规划路径的有效性；如果考虑攻击行为的不确定性，路径规划的适用性将更强，但是求解复杂度也将会更大。通常而言，不确定性的刻画需要独立重复大量的实验才能够较为准确地刻画攻击成功的概率，消耗大量时间。因此不确定性的精确度量是实现面向自动化渗透测试攻击路径发现的重要因素。

过程动态性：渗透测试是一个动静结合的过程，渗透测试中的静态层面包括目标主机系统架构，网络结构等在一次渗透测试过程中基本上是稳定不变的；而动态层面是指网络配置变化，主机更新补丁等。主机状态的动态性往往导致攻击结果的变化，例如同一个攻击载荷攻击目标主机时，当目标主机是否更新补丁将会造成不一样的攻击结果。渗透测试过程中的动态性往往伴随着攻防博弈过程，因此当不考虑渗透测试过程动态性时，渗透测试过程是一个单方面的攻击过程，这样会大大减少攻击规划的复杂度，但是会限制攻击规划的应用场景；而当考虑渗透测试过程动态性时，渗透测试过程变成一个动态博弈过程，攻击规划场景的适用性将会大大提高，但与此同时路径规划的复杂度也随之提升。

资源约束性：渗透测试过程中，攻防双方都存在资源、时间等约束条件。对攻击方而言，攻击手段、渗透测试时间、渗透测试代价的限制会严重影响攻击路径的选择。对防御方而言，随着自身网络规模的扩大，不可能完全考虑到每一台主机自身的安全设置问题，如何将有限的资源分配到自身网络关键设施上，保证自身网络能够正常、安全的运行对防御方而言是一种严峻的挑战。

路径最优性：在渗透测试过程中由于渗透测试人员安全技能差异、攻击动作代价和收益的不同，往往会存在多种攻击方案达到渗透目的。因此需要从多种可行的攻击方案中选择出效益-成本最优的攻击路径。而这种最优方案的选择不仅仅与攻击行为成功的概率有关，还与攻击者自身的技能熟练度，以及攻击动作代价等相关，因此需要综合考虑各方面因素进行最优攻击路径发现。

4.2 领域独立智能规划算法用于攻击路径发现的优缺点

如表1所示，本文具体分析了现有的领域独立智能规划算法，从其所属技术类型、观测完备性、行为不确定性、动态性、资源约束性和路径最优性5个方面进行了对比分析，并对每一类领域独立智能规划算法的优缺点进行了分析总结。整体而言，确定性规划算法中基于图规划技术的智能规划算法能够适用于较大的问题规模，且能够搜索得到最优路径，因此得到了广泛的应用。虽然基于HTN技术的规划算法适用范围更大，算法复杂度更低，但是由于其需要人工设置先验任务分解方法，从而会严重影响路径规划的自动化程度，因此不适用于自动化渗透测试场景下的攻击路径发现。非确定性规划相对于确定性规划算法在进行渗透测试路径规划时具有更好的适用性，能够刻画渗透测试的状态空间完备性、行为不确定性以及路径最优性，但计算复杂度较高，如何提高非确定性条件下的路径规划算法求解效率是该类算法未来的研究重点。博弈规划模型能够结合渗透测试过程中的攻防对抗特性进行攻击路径发现，提高发现攻击路径的有效性，但该类算法存在较强的模型假设，导致实际应用受限，提高模型适用性是该类方法的研究重点。

表 1　领域独立智能规划算法进行攻击路径发现时的适用性总结

	类型	文献	O	U	D	R	M	优点	缺点
确定性攻击路径发现	规划图	[18]	√	×	×	×	√	能够显示描述所有可能攻击路径，可解释性强	时间复杂度高，为$O(mn^k)$，m为状态空间大小，n为动作空间大小，k为层数
		[20]	√	×	×	×	√	基于规划图构建启发函数，提高攻击路径发现效率	时间复杂度高，为$O(m^n)$，不适用于大规模场景m为状态空间大小，n为动作空间大小
	偏序规划	[22]	√	×	×	×	×	能够发现所有动作对之间的约束关系	需要遍历动作空间，构建约束集合，造成额外时间开销
		[24]	√	×	×	×	√	构造启发函数选择动作，并利用约束关系缩减规模，提高路径搜索效率	
	分层任务网络	[30]	√	×	×	×	×	可解释性更强	需要专家制定分解方法
		[31]	√	×	×	×	√	利用标准优化算法提高路径发现效率	
	Determinizing	[36]	√	√	×	×	×	可扩展性好，适用多种非确定性场景	无法进行重规划
	概率优化	[41]	√	√	×	×	×	能够根据实际执行结果进行重规划	需要删除非确定性信息进行规划，无法利用规划反馈信息
		[44]	√	×	×	×	√	构造规划图启发函数，求解效率高	构建多个规划图，造成大量冗余
	马尔可夫决策过程	[52]	√	√	×	×	√	能存储大规模网络空间状态策略，策略求解效率更高	容易陷入局部极小值
		[53]	√	√	×	×	√	基于数据确定模型的参数个数和函数形式，无需人工设定，灵活方便	在较大数据集的情况下训练时间较长
非确定性攻击路径发现		[55]	√	√	×	×	√	精确求解算法，是后续近似求解算法的基础	求解复杂度极高，当状态空间较大时无法进行规划求解
		[57]	√	√	×	×	√	首个基于点迭代的近似求解方法，求解效率相对于精确求解效率高	仅能对单主机进行规划，时间复杂度$O(\|N\|\|A\|(\|S\|\|B\|+\|O\|))$，其中$S$为状态集合，$A$为动作集合，$O$为观测状态集合，$B$为信念状态点集合，$N$为上限点集合
	部分观测的马尔可夫决策过程	[58]	√	√	×	×	√	采用前向搜索策略，采样效率更高，适合短序列场景	仅能对单主机进行规划，时间复杂度$O(\|N\|(\|S\|^2+\|A\|+\|O\|))$其中$S$为状态集合，$A$为动作集合，$O$为观测状态集合，$N$为上限点集合
		[59]	√	√	×	×	√	采样效率高	仅能对单主机进行规划，无法扩展到网络层面，时间复杂度为$O(\|S\|^3\|A\|\|O\|\|B\|\|N\|)$其中$S$为状态集合，$A$为动作集合，$O$为观测状态集合，$B$为信念状态点集合，$N$为上限点集合
		[60]	√	√	×	×	√	能够实现网络层面攻击路径发现	假定网络拓扑结构及策略稳定不变
博弈攻击路径发现	静态博弈模型	[62]	√	×	×	×	√	首次将博弈模型引入到攻防对抗	要求完全信息且攻防双方为完全理性，并且要求攻防对抗策略保持不变
		[63]	√	×	×	×	√	求解效率高	
	动态博弈模型	[67]	√	×	√	×	√	多轮次博弈条件下的攻击路径发现	要求完全信息且攻防双方为完全理性
		[68]	√	×	√	×	√	摒弃了完全理性和完全信息假设	复杂度较高，为$O((m+n)^2)$，m和n分别为攻防策略集合大小
		[71]	√	×	√	×	√	摒弃了攻防双方对等信息的假设	模型复杂，求解难，现实应用场景受限

注：O：状态空间完备性；U：行为不确定性；D:过程动态性；R：资源约束性；M：路径最优性。

5　未来研究方向

5.1　大规模网络场景下的快速攻击路径发现

现有的攻击路径发现算法研究由于其计算复杂度较高，只适用于小规模网络场景，无法对大规模网络场景进行有效快速攻击路径发现，因此研究大规模网络场景下的快速攻击路径发现技术具有重要意义。现有的规划算法在进行规划过程中会产生大量无效操作，例如规划图算法会扩展生成大量与目标状态无关的中间状态、偏序规划会存储大量无用约束对等。以目标状态为牵引，限制规划算法状态无限扩张，进行有效剪枝操作是解决该问题的有效途径。

5.2　结合反馈信息的攻击路径发现

现有的攻击路径发现算法研究以单次规划为

主，未考虑到规划路径有效性，而实际测试证明单次规划出来的攻击路径往往不能够成功达到测试目的。产生该现象的原因是由于先验知识的缺失导致无法准确定位有效攻击载荷，进而导致攻击失效的现象。将渗透测试过程中产生的反馈信息融入到攻击路径发现算法当中动态调整攻击路径是解决该问题的有效途径。反馈信息是对攻击载荷以及主机状态信息的有效反映，将渗透测试过程中产生的反馈信息融入到路径规划算法当中能够有效剔除无效攻击载荷，提高规划路径有效性。例如根据反馈信息动态修改载荷代价，降低有效攻击载荷行为代价，提高无效载荷行为代价，引导启发函数最小化规划代价，从而达到提高规划路径有效性的目的。

5.3 多目标环境下攻击路径发现

现有算法着重研究单目标条件下的攻击路径发现，渗透测试目标唯一，但实际渗透测试需要达到多种目标。在多目标环境下，现有算法将会产生大量冗余局部路径，造成较大的时间开销和计算资源的浪费。因此协同调整资源利用，实现多目标环境下最小代价最大成功率攻击路径发现成为亟需解决的问题。多无人机协同路径规划等相关研究为解决该问题提供了解决思路。

6 结束语

攻击路径发现是自动化渗透测试领域的研究热点。本文首先论述了领域相关攻击路径发现算法应用于自动化渗透测试存在的缺陷，指出领域独立智能规划算法用于自动化渗透测试条件下攻击路径发现的必要性。其次对领域独立智能规划技术研究进展进行总结，介绍了确定性规划算法、非确定性规划算法以及博弈规划模型的原理，然后系统地梳理了领域独立智能规划技术应用于面向自动化渗透测试的攻击路径发现问题的技术特点及适用性，最后分析当前面向自动化渗透测试的攻击路径发现面临的挑战，以期对未来研究提供启发。

参 考 文 献

[1] KRUTZ R L and VINES R D. The CISSP and CAP Prep guide: Platinum Edition[M]. New Jersey: Wiley, 2007.

[2] STEFINKO Y, PISKOZUB A, and BANAKH R. Manual and automated penetration testing. Benefits and drawbacks. Modern tendency[C]. The 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science, Lviv, Ukraine, 2016: 488–491. doi: 10.1109/tcset.2016.7452095.

[3] ABU-DABASEH F and ALSHAMMARI E. Automated penetration testing: An overview[C]. The 4th International Conference on Natural Language Computing, Copenhagen, Denmark, 2018: 121–129.

[4] MCDERMOTT J P. Attack net penetration testing[C]. 2000 Workshop on New Security Paradigms, Ballycotton, Ireland, 2001: 15–21. doi: 10.1145/366173.366183.

[5] 诸葛建伟, 陈力波, 孙松柏, 等. Metasploit渗透测试魔鬼训练营[M]. 北京: 机械工业出版社, 2013: 3–4.
ZHUGE Jianwei, CHEN Libo, SUN Songbai, *et al*. Penetration Testing Devil Training Camp Based on Metasploit[M]. Beijing: China Machine Press, 2013: 3–4.

[6] POLATIDIS N, PAVLIDIS M, and MOURATIDIS H. Cyber-attack path discovery in a dynamic supply chain maritime risk management system[J]. *Computer Standards & Interfaces*, 2018, 56: 74–82. doi: 10.1016/j.csi.2017.09.006.

[7] 李庆华, 尤越, 沐雅琪, 等. 一种针对大型凹型障碍物的组合导航算法[J]. 电子与信息学报, 2020, 42(4): 917–923. doi: 10.11999/JEIT190179.
LI Qinghua, YOU Yue, MU Yaqi, *et al*. Integrated navigation algorithm for large concave obstacles[J]. *Journal of Electronics & Information Technology*, 2020, 42(4): 917–923. doi: 10.11999/JEIT190179.

[8] BIALEK Ł, DUNIN-KĘPLICZ B, and SZAŁAS A. A paraconsistent approach to actions in informationally complex environments[J]. *Annals of Mathematics and Artificial Intelligence*, 2019, 86(4): 231–255. doi: 10.1007/s10472-019-09627-9.

[9] AMMANN P, WIJESEKERA D, and KAUSHIK S. Scalable, Graph-based network vulnerability analysis[C]. The 9th ACM Conference on Computer and Communications Security, Washington, USA, 2002: 217–224. doi: 10.1145/586110.586140.

[10] CHEN Feng, LIU Dehui, ZHANG Yi, *et al*. A scalable approach to analyzing network security using compact attack graphs[J]. *Journal of Networks*, 2010, 5(5): 543–550. doi: 10.4304/jnw.5.5.543-550.

[11] OU Xinming, GOVINDAVAJHALA S, and APPEL A W. MulVAL: A logic-based network security analyzer[C]. The 14th Conference on USENIX Security Symposium, Baltimore, USA, 2005: 113–128.

[12] WANG Lingyu, YAO Chao, SINGHAL A, *et al*. Interactive analysis of attack graphs using relational queries[C]. The 20th Annual Conference on Data and Applications Security and Privacy, Sophia Antipolis, France, 2006: 119–132. doi: 10.1007/11805588_9.

[13] LI Wei, VAUGHN R B, and DANDASS Y S. An approach to model network exploitations using exploitation graphs[J] *Simulation*, 2006, 82(8): 523–541. doi: 10.1177/0037549706072046.

[14] MAHDAVI A and CARVALHO M. Optimal trajectory and schedule planning for autonomous guided vehicles in flexible manufacturing system[C]. The 2nd IEEE International

Conference on Robotic Computing, Laguna Hills, USA, 2018: 167–172. doi: 10.1109/irc.2018.00034.

[15] MA Xiaobai, JIAO Ziyuan, WANG Zhenkai, *et al*. 3-D decentralized prioritized motion planning and coordination for high-density operations of micro aerial vehicles[J]. *IEEE Transactions on Control Systems Technology*, 2018, 26(3): 939–953. doi: 10.1109/tcst.2017.2699165.

[16] ZANG Yichao, ZHOU Tianyang, GE Xiaoyue, *et al*. An improved attack path discovery algorithm through compact graph planning[J]. *IEEE Access*, 2019, 7: 59346–59356. doi: 10.1109/access.2019.2915091.

[17] BODDY M S, GOHDE J, HAIGH T, *et al*. Course of action generation for cyber security using classical planning[C]. The 15th International Conference on Automated Planning and Scheduling, Monterey, USA, 2005: 12–21.

[18] GARRETT C R, LOZANO-PÉREZ T, and KAELBLING L P. FFRob: Leveraging symbolic planning for efficient task and motion planning[J]. *The International Journal of Robotics Research*, 2018, 37(1): 104–136. doi: 10.1177/0278364917739114.

[19] KAUTZ H A and SELMAN B. Unifying SAT-based and graph-based planning[C]. The 16th International Joint Conference on Artificial Intelligence, Stockholm, Sweden, 1999: 318–325.

[20] DO M B and KAMBHAMPATI S. Planning as constraint satisfaction: Solving the planning graph by compiling it into CSP[J]. *Artificial Intelligence*, 2001, 132(2): 151–182. doi: 10.1016/s0004-3702(01)00128-x.

[21] BAIOLETTI M, MARCUGINI S, and MILANI A. DPPlan: An algorithm for fast solutions extraction from a planning graph[C]. The 5th International Conference on Artificial Intelligence Planning Systems, Breckenridge, USA, 2000: 13–21.

[22] BARRETT A and WELD D S. Partial-order planning: Evaluating possible efficiency gains[J]. *Artificial Intelligence*, 1994, 67(1): 71–112. doi: 10.1016/0004-3702(94)90012-4.

[23] NGUYEN X L and KAMBHAMPATI S. Reviving partial order planning[C]. The 17th International Joint Conference on Artificial Intelligence, Seattle, USA. 2001: 459–466.

[24] YOUNES H L S and SIMMONS R G. VHPOP: Versatile heuristic partial order planner[J]. *Journal of Artificial Intelligence Research*, 2003, 20: 405–430. doi: 10.1613/jair.1136.

[25] COLES A J, COLES A, FOX M, *et al*. Forward-chaining partial-order planning[C]. The 20th International Conference on Automated Planning and Scheduling, Toronto, Canada, 2010: 42–49.

[26] BOUTILIER C and BRAFMAN R I. Partial-order planning with concurrent interacting actions[J]. *Journal of Artificial Intelligence Research*, 2001, 14: 105–136. doi: 10.1613/jair.740.

[27] MOHR F, WEVER M, and HÜLLERMEIER E. ML-Plan: Automated machine learning via hierarchical planning[J]. *Machine Learning*, 2018, 107(8–10): 1495–1515. doi: 10.1007/s10994-018-5735-z.

[28] DE SILVA L, PADGHAM L, and SARDINA S. HTN-like solutions for classical planning problems: An application to BDI agent systems[J]. *Theoretical Computer Science*, 2019, 763: 12–37. doi: 10.1016/j.tcs.2019.01.034.

[29] SOHN S, OH J, and LEE H. Hierarchical reinforcement learning for zero-shot generalization with subtask dependencies[C]. The 32nd Conference on Neural Information Processing Systems, Montréal, Canada, 2018: 7156–7166.

[30] MU Chengpo and LI Yingjiu. An intrusion response decision-making model based on hierarchical task network planning[J]. *Expert Systems with Applications*, 2010, 37(3): 2465–2472. doi: 10.1016/j.eswa.2009.07.079.

[31] ONTAÑÓN S and BURO M. Adversarial hierarchical-task network planning for complex real-time games[C]. The 24th International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 2015: 1652–1658.

[32] FU JICHENG, NG V, BASTANI F B, *et al*. Simple and fast strong cyclic planning for fully-observable nondeterministic planning problems[C]. The 22nd International Joint Conference on Artificial Intelligence, Barcelona, Spain, 2011: 1949–1954. doi: 10.1007/s10472-016-9517-7.

[33] KOLOBOV A, MAUSAM M, and Weld D S. LRTDP versus UCT for online probabilistic planning[C]. The 26th AAAI Conference on Artificial Intelligence, Toronto, Canada, 2012: 1786–1792.

[34] YOON S, FERN A, GIVAN R, *et al*. Probabilistic planning via determinization in hindsight[C]. The 23rd AAAI Conference on Artificial Intelligence, Chicago, USA, 2008: 1010–1016.

[35] CIMATTI A, PISTORE M, ROVERI M, *et al*. Weak, strong, and strong cyclic planning via symbolic model checking[J]. *Artificial Intelligence*, 2003, 147(1/2): 35–84. doi: 10.1016/s0004-3702(02)00374-0.

[36] MUISE C J, MCILRAITH S A, and BECK J C. Improved non-deterministic planning by exploiting state relevance[C]. The 22nd International Conference on Automated Planning and Scheduling, Atibaia, Brazil, 2012: 172–180.

[37] MUISE C J, MCILRAITH S A, and BELLE V. Non-deterministic planning with conditional effects[C]. The 24th International Conference on Automated Planning and Scheduling, Portsmouth, USA, 2014: 370–374.

[38] 李洋, 文中华, 伍小辉, 等. 求最小期望权值强循环规划解[J]. 计算机科学, 2015, 42(4): 217–220, 257. doi: 10.11896/j.issn.

1002-137X.2015.04.044.

LI Yang, WEN Zhonghua, WU Xiaohui, *et al*. Solving strong cyclic planning with minimal expectation weight[J]. *Computer Science*, 2015, 42(4): 217–220, 257. doi: 10.11896/j.issn.1002-137X.2015.04.044.

[39] 唐杰, 文中华, 汪泉, 等. 不确定可逆规划的强循环规划解[J]. 计算机研究与发展, 2013, 50(9): 1970–1980. doi: 10.7544/issn1000-1239.2013.20130404.

TANG Jie, WEN Zhonghua, WANG Quan, *et al*. Solving strong cyclic planning in nondeterministic reversible planning domain[J]. *Journal of Computer Research and Development*, 2013, 50(9): 1970–1980. doi: 10.7544/issn1000-1239.2013.20130404.

[40] KUSHMERICK N, HANKS S, and WELD D S. An algorithm for probabilistic planning[J]. *Artificial Intelligence*, 1995, 76(1/2): 239–286. doi: 10.1016/0004-3702(94)00087-H.

[41] YOON S W, FERN A, and GIVAN R. FF-Replan: A baseline for probabilistic planning[C]. The 17th International Conference on Automated Planning and Scheduling, Providence, USA, 2007: 352–359.

[42] YOON S, RUML W, BENTON J, *et al*. Improving determinization in hindsight for on-line probabilistic planning[C]. The 20th International Conference on Automated Planning and Scheduling, Toronto, Canada, 2010: 209–216.

[43] ISSAKKIMUTHU M, FERN A, KHARDON R, *et al*. Hindsight optimization for probabilistic planning with factored actions[C]. The 25th International Conference on Automated Planning and Scheduling, Jerusalem, Israel, 2015: 120–128.

[44] BRYCE D, KAMBHAMPATI S, and SMITH D E. Sequential Monte Carlo in probabilistic planning reachability heuristics[C]. The 16th International Conference on Automated Planning and Scheduling, Cumbria, UK, 2006: 233–242.

[45] TREVIZAN F W, THIÉBAUX S, and HASLUM P. Occupation measure heuristics for probabilistic planning[C]. The 27th International Conference on Automated Planning and Scheduling, Pittsburgh, USA, 2017: 306–315.

[46] DURKOTA K and LISÝ V. Computing optimal policies for attack graphs with action failures and costs[C]. The 7th European Starting AI Researcher Symposium, Prague, Czech Republic, 2014: 101–110.

[47] SUN Wen, GORDON G J, BOOTS B, *et al*. Dual policy iteration[C]. The 32nd Conference on Neural Information Processing Systems, Montréal, Canada, 2018: 7059–7069.

[48] HOUTHOOFT R, CHEN R Y, ISOLA P, *et al*. Evolved policy gradients[C]. The 32nd International Conference on Neural Information Processing Systems, Montréal, Canada, 2018: 5405–5414.

[49] LIU Huaping, WU Yupei, and SUN Fuchun. Extreme trust region policy optimization for active object recognition[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, 29(6): 2253–2258. doi: 10.1109/TNNLS.2017.2785233.

[50] SRINIVASAN S, LANCTOT M, ZAMBALDI V, *et al*. Actor-critic policy optimization in partially observable multiagent environments[C]. The 32nd Conference on Neural Information Processing Systems, Montréal, Canada, 2018: 3422–3435.

[51] KANG Qinma, ZHOU Huizhuo, and KANG Yunfan. An asynchronous advantage actor-critic reinforcement learning method for stock selection and portfolio management[C]. The 2nd International Conference on Big Data Research, Weihai, China, 2018: 141–145. doi: 10.1145/3291801.3291831.

[52] TAN Fuxiao and GUAN Xinping. Kernel-based adaptive critic designs for optimal control of nonlinear discrete-time system[C]. The 37th Chinese Control Conference, Wuhan, China, 2018: 2167–2172. doi: 10.23919/chicc.2018.8482778.

[53] TAYLOR G and PARR R. Kernelized value function approximation for reinforcement learning[C]. The 26th Annual International Conference on Machine Learning, Montreal, Canada, 2009: 1017–1024. doi: 10.1145/1553374.1553504.

[54] SARRAUTE C, BUFFET O, and HOFFMANN J. Penetration testing== POMDP solving?[C]. 2011 IJCAI Workshop on Intelligent Security, Barcelona, Spain, 2011: 66–73.

[55] SMALLWOOD R D and SONDIK E J. The optimal control of partially observable Markov processes over a finite horizon[J]. *Operations Research*, 1973, 21(5): 1071–1088. doi: 10.1287/opre.21.5.1071.

[56] CHENG H T. Algorithms for partially observable Markov decision processes[D]. [Ph. D. dissertation], The University of British Columbia, 1988. doi: 10.14288/1.0098252.

[57] PINEAU J, GORDON G, and THRUN S. Point-based value iteration: An anytime algorithm for POMDPs[C]. The 18th International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 2003: 1025–1032.

[58] LIU Bingbing, KANG Yu, JIANG Xiaofeng, *et al*. A fast approximation method for partially observable Markov decision processes[J]. *Journal of Systems Science and Complexity*, 2018, 31(6): 1423–1436. doi: 10.1007/s11424-018-7038-7.

[59] KURNIAWATI H, HSU D, LEE W S. SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces[C]. The Robotics: Science and Systems IV, Zurich, Switzerland, 2008: 65–72. doi:

10.15607/RSS.2008.IV.009.

[60] SARRAUTE C, BUFFET O, and HOFFMANN J. POMDPs make better hackers: Accounting for uncertainty in penetration testing[C]. The 26th AAAI Conference on Artificial Intelligence, Toronto, Canada, 2012: 1816–1824.

[61] 王刚, 胡鑫, 马润年, 等. 集体防御机制下的网络行动同步建模和稳定性[J]. 电子与信息学报, 2018, 40(6): 1515–1519. doi: 10.11999/JEIT170619.
WANG Gang, HU Xin, MA Runnian, *et al*. Synchronization modeling and stability of cyberspace operation based on collective defensive mechanism[J]. *Journal of Electronics & Information Technology*, 2018, 40(6): 1515–1519. doi: 10.11999/JEIT170619.

[62] LYE K W and WING J M. Game strategies in network security[J]. *International Journal of Information Security*, 2005, 4(1/2): 71–86. doi: 10.1007/s10207-004-0060-x.

[63] 姜伟, 方滨兴, 田志宏, 等. 基于攻防博弈模型的网络安全测评和最优主动防御[J]. 计算机学报, 2009, 32(4): 817–827. doi: 10.3724/SP.J.1016.2009.00817.
JIANG Wei, FANG Binxing, TIAN Zhihong, *et al*. Evaluating network security and optimal active defense based on attack-defense game model[J]. *Chinese Journal of Computers*, 2009, 32(4): 817–827. doi: 10.3724/SP.J.1016.2009.00817.

[64] 王晋东, 余定坤, 张恒巍, 等. 静态贝叶斯博弈主动防御策略选取方法[J]. 西安电子科技大学学报: 自然科学版, 2016, 43(1): 144–150. doi: 10.3969/j.issn.1001-2400.2016.01.026.
WANG Jindong, YU Dingkun, ZHANG Hengwei, *et al*. Active defense strategy selection based on the static Bayesian game[J]. *Journal of Xidian University*, 2016, 43(1): 144–150. doi: 10.3969/j.issn.1001-2400.2016.01.026.

[65] 王元卓, 林闯, 程学旗, 等. 基于随机博弈模型的网络攻防量化分析方法[J]. 计算机学报, 2010, 33(9): 1748–1762. doi: 10.3724/SP.J.1016.2010.01748.
WANG Yuanzhuo, LIN Chuang, CHENG Xueqi, *et al*. Analysis for network attack-defense based on stochastic game model[J]. *Chinese Journal of Computers*, 2010, 33(9): 1748–1762. doi: 10.3724/SP.J.1016.2010.01748.

[66] CUI Xiaolin, TAN Xiaobin, ZHANG Yong, *et al*. A Markov game theory-based risk assessment model for network information system[C]. 2008 International Conference on Computer Science and Software Engineering, Hubei, China, 2008: 1057–1061. doi: 10.1109/csse.2008.949.

[67] LI Tao, WANG Jindong, CHEN Yu, *et al*. A multi-stage game approach applied to network security risk controlling[C]. The 2nd IEEE Advanced Information Technology, Electronic and Automation Control Conference, Chongqing, China, 2017: 2518–2522. doi: 10.1109/iaeac.2017.8054477.

[68] 黄健明, 张恒巍, 王晋东, 等. 基于攻防演化博弈模型的防御策略选取方法[J]. 通信学报, 2017, 38(1): 168–176. doi: 10.11959/j.issn.1000-436x.2017019.
HUANG Jianming, ZHANG Hengwei, Wang Jindong, *et al*. Defense strategies selection based on attack-defense evolutionary game model[J]. *Journal on Communications*, 2017, 38(1): 168–176. doi: 10.11959/j.issn.1000-436x.2017019.

[69] 张恒巍, 李涛. 基于多阶段攻防信号博弈的最优主动防御[J]. 电子学报, 2017, 45(2): 431–439. doi: 10.3969/j.issn.0372-2112.2017.02.023.
ZHANG Hengwei and LI Tao. Optimal active defense based on multi-stage attack-defense signaling game[J]. *Acta Electronica Sinica*, 2017, 45(2): 431–439. doi: 10.3969/j.issn.0372-2112.2017.02.023.

[70] 朱建明, 宋彪, 黄启发. 基于系统动力学的网络安全攻防演化博弈模型[J]. 通信学报, 2014, 35(1): 54–61. doi: 10.3969/j.issn.1000-436x.2014.01.007.
ZHU Jianming, SONG Biao, and HUANG Qifa. Evolution game model of offense-defense for network security based on system dynamics[J]. *Journal on Communications*, 2014, 35(1): 54–61. doi: 10.3969/j.issn.1000-436x.2014.01.007.

[71] VADLAMUDI S G, SENGUPTA S, TAGUINOD M, *et al*. Moving target defense for web applications using Bayesian stackelberg games: (Extended Abstract)[C]. The 2016 International Conference on Autonomous Agents & Multiagent Systems, Singapore, 2016: 1377–1378.

[72] YUAN Yuan, SUN Fuchun, and LIU Huaping. Resilient control of cyber-physical systems against intelligent attacker: A hierarchal stackelberg game approach[J]. *International Journal of Systems Science*, 2016, 47(9): 2067–2077. doi: 10.1080/00207721.2014.973467.

臧艺超：男，1991年生，博士生，研究方向为路径规划，强化学习，效果评估.

周天阳：男，1979年生，副教授，研究方向为网络安全，强化学习，效果评估.

朱俊虎：男，1971年生，教授，研究方向为网络安全，网络模拟与效果评估.

王清贤：男，1960年生，教授，研究方向为网络安全，计算复杂度，网络模拟与效果评估.

责任编辑：陈　倩