计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202330961
61(5): 1230−1249, 2024

欺骗防御技术发展及其大语言模型应用探索

王 瑞 阳长江 邓向东 刘 园 田志宏
(广州大学网络空间安全学院 广州 510799)
(ruiwang@e.gzhu.edu.cn)

Development of Deception Defense Technology and Exploration of Its Large Language Model Applications

Wang Rui, Yang Changjiang, Deng Xiangdong, Liu Yuan, and Tian Zhihong
(*Cyberspace Institute of Advanced Technology*, *Guangzhou University*, *Guangzhou* 510799)

Abstract Deception defense, as the most promising technology in proactive defense, aids defenders in facing highly covert and unknown threats, turning passivity into proactivity, and breaking the inherent imbalance between offense and defense. In the face of potential threat scenarios, how to effectively use deception defense technology to help defenders anticipate threats, perceive threats, and entrap threats, is a key issue that currently need to be addressed. Game theory and attack graph models provide strong support in formulating active defense strategies and analyzing potential risks. We summarize and review the recent work of both in the realm of deception defense. With the rapid development of large language model technology and its increasingly close integration with the field of cybersecurity, we review traditional deception defense technology and propose a large language model-based intelligent external network HoneyPoint generation technique. Experimental analysis validates the effectiveness of external network HoneyPoint in capturing network threats, showing improvements over traditional Web honeypots in aspects like simulation, stability, and flexibility. To enhance the collaborative cooperation between HoneyPoints and improve the capabilities for threatening exploration and perception, the concept of Honey-Landscape is introduced. We provide an outlook on how to utilize HoneyPoint and Honey-Landscape technologies to construct an integrated active defense mechanism that includes threat prediction, threat perception, and threat entrapment.

Key words deception defense; large language model; attack graph; game theory; HoneyPoint; Honey-Landscape

摘 要 欺骗防御作为主动防御中最具发展前景的技术,帮助防御者面对高隐蔽未知威胁化被动为主动,打破攻守间天然存在的不平衡局面.面对潜在的威胁场景,如何利用欺骗防御技术有效地帮助防御者做到预知威胁、感知威胁、诱捕威胁,均为目前需要解决的关键问题.博弈理论与攻击图模型在主动防御策略制定、潜在风险分析等方面提供了有力支撑,总结回顾了近年来二者在欺骗防御中的相关工作.随着大模型技术的快速发展,大模型与网络安全领域的结合也愈加紧密,通过对传统欺骗防御技术的回顾,提出了一种基于大模型的智能化外网蜜点生成技术,实验分析验证了外网蜜点捕获网络威胁的有效性,与传统 Web 蜜罐相比较,在仿真性、稳定性与灵活性等方面均有所提升.为增强蜜点间协同合作、提升对攻击威胁的探查与感知能力,提出蜜阵的概念.针对如何利用蜜点和蜜阵技术,对构建集威胁预测、威胁感知和威胁诱捕为一体的主动防御机制进行了展望.

收稿日期: 2023-11-30; **修回日期**: 2024-03-18
基金项目: 国家自然科学基金项目 (U20B2046); 国家重点研发计划项目 (2021YFB2012402); 广东省高校珠江学者资助计划 (2019)
This work was supported by the National Natural Science Foundation of China (U20B2046), the National Key Research and Development Program of China (2021YFB2012402), and Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme(2019).
通信作者: 田志宏(tianzhihong@gzhu.edu.cn)

关键词 欺骗防御；大语言模型；攻击图；博弈论；蜜点；蜜阵

中图法分类号 TP391

“兵者，诡道也”，在战争中合理地运用欺骗性战术往往能够取得出奇制胜的效果[1]. 面对日益猖獗、无孔不入的网络攻击，以注重边界防御为核心的防火墙、访问控制、反病毒软件、入侵检测系统（IDS）和入侵防御系统（IPS）等均已被证明无法防御渗透等高级威胁[2]. 尤其是面对高级持续性威胁[3]（APT），攻击者往往对目标系统进行持久性的网络侦察与漏洞信息收集[4]，利用精心设计的钓鱼邮件[5]等社会工程学手段寻找突破口，甚至通过零日漏洞[6]进行精准打击，建立立足点，在内部网络中横向移动[7]收集更多信息，最终实现破坏关键基础设施或窃取隐私数据的目的. 面对隐秘的攻击者，如何打破被动不利的局面是长久以来困扰防御者的关键难题，而欺骗性防御[8]技术的出现为打破僵局提供了契机.

欺骗防御（deception defense），通过混淆、误导、扭曲、操纵攻击者的信念，消磨攻击者的进攻欲望，浪费攻击者的进攻成本，模糊攻击者的进攻路径，扰乱攻击者的进攻目标，从而做到“不战而屈人之兵”[9]. 随着攻防双方的技术不断迭代更新，以欺骗防御为核心的主动防御技术得到了快速地发展. 早期以蜜罐（Honeypot）为核心的攻击诱捕体系，利用低交互蜜罐吸引、诱捕攻击者，对攻击行为进行初步分析[10]；使用高交互蜜罐与攻击者进行深层次地交互，收集并分析攻击者的攻击手段与策略[11]；利用多个蜜罐组成蜜网，模拟更加真实的网络环境，以此吸引、粘滞攻击，捕获攻击者的 TTPs[12]. 近年来 SDN、云计算技术的发展，轻量级、虚拟化蜜罐的出现使其部署方式更加灵活、模拟服务的类型愈加丰富，该阶段蜜罐捕获非定向性的攻击能力有显著的提升[13].

“上兵伐谋，其次伐交，其次伐兵，其下攻城”，为更有效地从欺骗策略层面对抗攻击者，研究者们将博弈论[14-15]、人工智能[16-17]、机器学习[18-20]、心理学[21-22]等多领域的先进技术和理论与欺骗理念相结合，使得欺骗防御进入到自动化与智能化决策的崭新阶段. 如图 1 所示，我们总结了 Web of Science 数据库中近年来相关文献的关键词，其中网络欺骗（cyber deception）、博弈（game theory）、防御策略（defense strategy）等均为目前研究的热点.

如表 1 所示，目前已有多篇针对欺骗防御发展的综述工作[23-26]，研究者们已从技术发展层面、攻防博弈演进等角度进行了总结和归纳. “知己知彼，百战不殆”，只有充分地了解敌手，从攻击者的角度对自

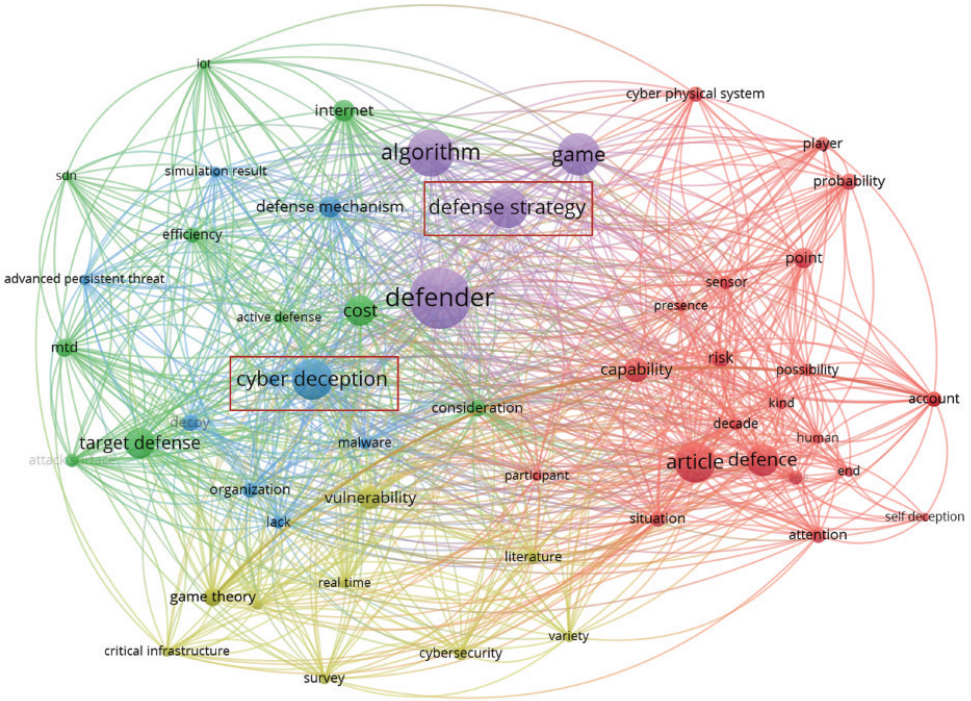


Fig. 1 Keywords of deception defense-related literature

图 1 欺骗防御相关文献关键词

Table 1 Existing Deception Defense-Related Review Work

表 1 现有欺骗防御相关综述工作

文献来源	研究角度	贡献
文献 [23]	网络欺骗形式化	对网络欺骗进行了形式化定义，概述了网络欺骗发展历程的 3 个阶段，将网络欺骗与网络杀伤链结合，提出了网络欺骗层次化模型，分析了网络欺骗在设备层、网络层、数据层、应用层的欺骗技术，并在网络杀伤链上进行了验证性的讨论.
文献 [24]	蜜罐（Honeypot）、蜜标（Honeytoken）、移动目标防御（MTD）	对近 30 年蜜罐、蜜标以及移动目标防御中代表性技术的整理，描述了 3 个领域之间关键技术的相互补充，并构建了基于欺骗的主动防御体系. 提出一个全新的杀伤链模型，从攻击阶段与欺骗层次两方面对 3 种主动防御技术进行了归类.
文献 [25]	博弈论	从博弈论的角度对欺骗防御的相关研究成果进行了筛选，提出了欺骗博弈的概念，并给出了网络欺骗博弈的形式化定义.
文献 [26]	博弈论、机器学习	从防御者的角度出发，在博弈论与机器学习两方面对防御性欺骗工作进行了较为全面地调查，阐述了防御性欺骗的设计原则与特性，明确了如何选取欺骗攻击者的类型、欺骗发起的时机以及欺骗技术的运用.

身的网络系统进行准确地风险评估，识别潜在的威胁，搜寻可能被利用的攻击路径，确定系统的漏洞分布以及需要修复的优先级，在上述信息的基础上才能构建更贴近现实场景的攻防博弈模型，从而使得欺骗策略有的放矢. 攻击图作为能够帮助防御者实现上述目标的强有力的工具，虽然目前已存在利用攻击图指导欺骗防御的相关工作，但关于如何将攻击图、博弈论与欺骗防御相结合，建立一个面向威胁探查的主动防御框架还没有相关的系统性调查. 因此，本文将从攻击图与博弈论两方面出发，对近年来欺骗防御技术相关工作展开调查.

此外，随着近几年大语言模型[27](large language model, LLM)的重要突破，尤其是 ChatGPT[28] 发布以后，其远超出传统算法的能力，极大地推动了人工智能和自然语言处理[29] 技术的发展，尤其是在语言理解、生成等方面展现出前所未有的能力，这也为欺骗防御技术的拓展开辟了新的可能性. 如何将大语言模型技术的优势辐射到欺骗防御领域成为了一个热点话题.

基于上述讨论可以发现，面对高隐蔽未知威胁，防御者在传统的网络安全防护模式下常常处于被动挨打、事后分析的窘迫处境. 欺骗防御作为最具发展前景的主动防御技术成为了打破僵局的重要利器. 本文旨在从攻击图与博弈论两方面出发，对欺骗防御技术展开调查，围绕着“以攻击图为依据，博弈理论为指导，欺骗防御为手段，大模型为辅助”的核心思想，研究并总结如何帮助防御者面对高隐蔽未知威胁，利用欺骗性手段将防御措施前置化，化被动为主动，反客为主，最终形成集威胁预测、威胁感知、威胁诱捕为一体的主动防御机制.

本文的主要贡献有 4 个方面：

1)从攻击图与博弈论相结合的角度，对欺骗防御技术相关研究展开调查并进行归纳总结.

2)从威胁模型出发，对欺骗防御技术进行了整体回顾，结合不同的攻击图模型与博弈论方法，深入分析了二者在威胁探查与智能决策中的作用.

3)针对传统蜜罐的局限性，探索大模型在欺骗防御领域中的应用，并提出了一种基于大模型的智能化外网蜜点(HoneyPoint)生成技术.

4)基于大模型、博弈论与攻击图模型在欺骗防御领域的应用，提出了“蜜阵”的概念，进而增强蜜点间的协同合作，提升对攻击威胁的探查与感知能力.

1 欺骗防御概述

1.1 欺骗防御定义

攻防博弈是一个动态演进的复杂过程，双方的技术在对抗中不断升级，其中欺骗策略在战场上的应用十分广泛. 军事上利用欺骗策略，可以隐藏、保护、加强、放大、最小化、掩盖己方的技术和意图. 在高度竞争、致命的环境中，在大规模作战行动中，欺骗可以成为实现作战突袭和保持主动性的关键因素. 其在网络攻防战中同样适用，防御者利用欺骗防御技术的目的是为了混淆、扰乱、迷惑攻击者，使其暴露自身信息，耗费更多的精力，从而延缓进攻或阻断进攻[30]. 下面给出欺骗防御的形式化定义：

定义 1. 欺骗防御(deception defense). 防御者利用欺骗性资源或手段，在被保护目标的网络或系统上布下陷阱或诱饵，实现扰乱、迷惑攻击进程，采集攻击活动的目的，使防御者提前感知威胁，变换防御部署，迫使攻击者花费更多的精力与成本，保证目标系统的安全性.

欺骗防御可由如下五元组进行形式化定义：

$$DD=(N,S,P,T,U).$$

1)$N=\{N_{\mathrm{A}},N_{\mathrm{D}}\}$表示欺骗防御实施阶段中的参与者，其中$N_{\mathrm{A}}$代表攻击者，其对目标网络发动攻击；$N_{\mathrm{D}}$代表防御者，是实施欺骗防御策略或行动的主体.

2）$S=\{S_A,S_D\}$表示欺骗防御过程中参与者的策略空间. $S_A=\{a_1,a_2,\cdots,a_m\},m\in\mathbb{N}^+$表示攻击者对被保护目标实施网络攻击时可采用的攻击策略，$S_D=\{d_1,d_2,\cdots,d_n\},n\in\mathbb{N}^+$表示防御者可使用的欺骗策略或可利用的防御性资源.

3）$P=\{p_1,p_2,\cdots,p_r\},r\in\mathbb{N}^+$表示被保护系统中的资产，如关键的主机和服务器或者隐秘的文件和数据等，其形式可以是任意网络中的关键组件.

4）$T:N\times S\times P\to\{0,1\}$表示部署欺骗防御资源后系统中的威胁检测函数. T=0 时，表示检测到威胁；T=1 时，表示检测到攻击.

5）$U:N\times S\times P\times T\to\mathbb{R}$，是防御者利用欺骗防御资源获取的收益. 其表现形式可以是攻击者所使用攻击技术、被保护目标的受损程度，以及消耗攻击者的攻击时间或精力等.

1.2 欺骗防御常用技术

1.2.1 蜜　罐

蜜罐作为欺骗防御体系中最基本、最核心的工具，其本质是通过模拟实际的计算机系统、网络服务和数据，诱使攻击者对其进行交互. 为了能够获取攻击者的关注，通常蜜罐会放在攻击者最易发现的路径上，并附加上较为明显的漏洞，增加暴露在攻击者视野下的概率. 一旦攻击者与蜜罐进行连接，便能够记录攻击者的行为信息，产生威胁预警. 蜜罐所具备的诱骗能力，其强弱取决于自身模拟的真实性与完整性的高低[31].

根据交互能力，蜜罐可分为高交互与低交互 2 种类型. 高交互的蜜罐能够模拟完整的操作系统和网络服务，提供与真实系统相似的交互体验，对于捕获攻击者的后续行为具有较强的能力，且与真实的环境越贴近，其被攻击者识破的几率越低. 而低交互蜜罐与高交互蜜罐相比较，其结构略微简单，通常仅模拟特定的服务或协议，但其维护成本将大大降低，在实际的部署应用时，低交互蜜罐可进行大量部署，增大攻击者触发蜜罐的可能[13].

根据部署位置，蜜罐可分为研究型与生产型 2 种类型. 研究型蜜罐一般为高交互蜜罐，其部署在公网用于收集黑客相关的行为和策略，通常由研究机构或政府组织进行维护和部署. 生产型蜜罐通常部署在内网，通过及时地与攻击者进行交互，发出攻击告警信息，帮助防御者尽早地掌握攻击者的行为动向[13].

1.2.2 蜜　网

不同于单个蜜罐，蜜网是由多个蜜罐和其他陷阱资源组成的网络环境，用来模拟真实的网络系统. 相比于蜜罐，蜜网的数据控制、数据捕获和数据收集能力更强大. 其突出的交互能力能够使攻击者意识不到自身正在蜜网的监视下，并且随着网络功能虚拟化（NFV）和软件定义网络（SDN）技术的成熟，蜜网可实现虚拟高交互蜜罐的自主部署和动态配置[32]. 多个蜜网的互联可形成蜜场，其在规模上更加庞大，适用于大规模的蜜罐管理和数据分析. 影子服务则比蜜网更有针对性，其通常针对被保护系统进行近似仿真，构造一个类似于目标系统的诱饵系统[33]. 蜜网等技术的本质是为了提升与攻击者的交互性，从而获得更多有利于防御者的威胁情报.

1.2.3 蜜　饵

蜜饵与蜜罐的核心目标具有一致性，均为吸引潜在的攻击者，但它们的实现方式有着显著的区别. 蜜饵根据模拟对象类型的不同和部署方式的差异，其存在方式多种多样. 蜜饵可以是虚假的登录凭据、虚假文档、虚假的数据库记录、虚假的 URL、虚假的配置文件、虚假电子邮件地址等[34]. 由于它们均有各自不同的特点，因此研究人员又将其中一些具有代表性的技术进行命名，如通常将与身份认证相关且具有溯源攻击者能力的技术称为蜜标（Honeytoken）等[35]. 由于上述技术依据的欺骗思路在本质上是相同的，为便于表述，本文中统一将其称为蜜饵.

1.2.4 移动目标防御（MTD）

绝大多数网络攻击都源于攻击者对目标的长期侦察，收集目标的相关信息，制作或利用针对性的网络武器进行精准打击. MTD 旨在通过不断变化系统的攻击面，以此来增加攻击者的攻击成本和复杂性. 其主要方法是将被保护系统的配置、软件、网络属性等进行动态变化，使得攻击者难以预测和发现有效攻击向量[36-38].

MTD 作为当前主流的主动防御方法之一，其与欺骗防御间的区别和包含关系一直以来存在着争议. 因为 MTD 中并不包含通过部署欺骗性的诱饵或虚假信息来误导攻击者的信念，仅是通过不断地变换来干扰和挫败攻击者的观测. 虽然 MTD 原本的概念中并不包含欺骗的思想，但是有效的欺骗防御技术往往需要配合动态部署和调整诱饵资源的位置与数量. 在实施欺骗的过程中，将 MTD 的相关手段和理念与之融合，往往能够获得更加理想的预期效果. 因此，在本文中将 MTD 视为欺骗防御的一部分进行相关工作的总结. 如图 2 所示，对上述欺骗防御相关技术进行了归纳与总结.

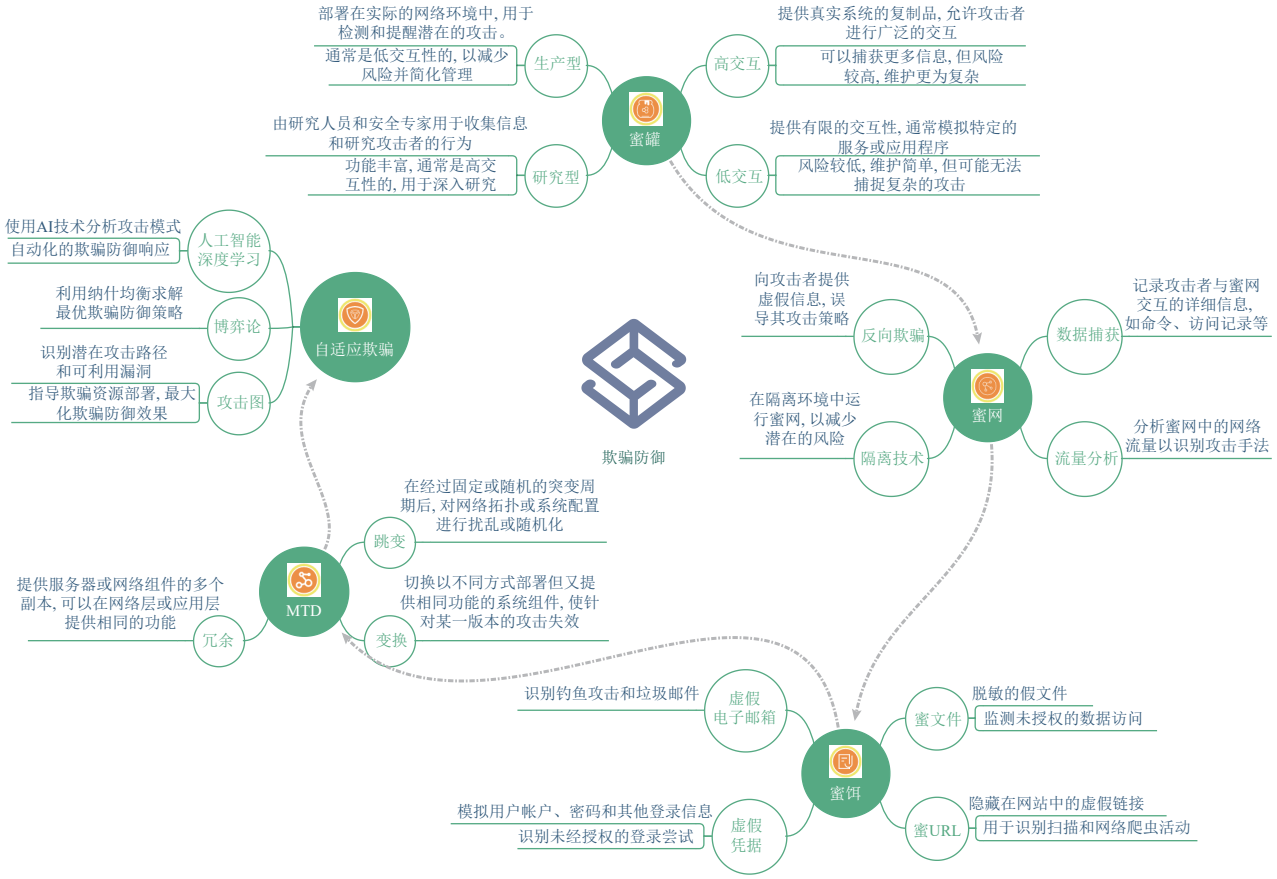


Fig. 2 Introduction of deception defense related technologies

图 2 欺骗防御相关技术介绍

2 威胁模型

在攻防对抗中攻击者往往占据着主动的优势，攻守间的不平衡性主要体现在：

1)信息不对称[39]. APT 攻击可能持续数月或数年，因此攻击者对目标系统的脆弱点、资产分布、防御措施等均进行了仔细地排查与了解，而防御者面对未知的敌手，需要时刻应对各类潜在的攻击手段.

2)成本不对称[40]. 攻击者实施成功的攻击仅需要找到系统中存在的一个脆弱点，而防御者则需要部署大量的防御成本，保护所有可能的入侵点，漫无目的的防护造成了防御资源的极大浪费.

3)时间不对称[41]. 面对攻击者采用零日漏洞等高威胁手段进攻，防御者常常处于后知后觉的被动局面，防御措施总是滞后于攻击，如何提前预知威胁、感知威胁也一直困扰着网络管理者.

为更好地应对时刻发生的网络攻击，理解攻击者的行为和意图，安全研究人员对网络攻击进行了高度的凝练和概括. 本节从攻击者视角和防御者视角 2 方面，分别对当前较为通用的攻击威胁与防御对抗模型进行了梳理，如表 2 所示.

2.1 攻击者视角

从攻击者视角对网络攻击过程进行建模，对于提高防御效率和理解安全威胁是至关重要的. 其通常具备 3 个优势：

1)更好地理解攻击者的行为和意图，通过模拟攻击者的策略，可以揭示攻击者如何选择目标、使用工具和执行攻击；

2)分析攻击者的行为模式有助于预测未来的攻击趋势，可以制定预防性安全策略，提前阻止攻击；

3)攻击者视角可以帮助识别网络和系统中的安全漏洞，指导防御者如何加固防御措施和减少攻击面.

洛克希德·马丁公司提出网络杀伤链[42](cyber kill chain)的概念，将网络攻击分为了 7 个阶段. Meta 采用与 Cyber Kill Chain 类似的方法，针对在线威胁，提出了“十阶段在线操作杀伤链[43]”，每个阶段都表示一个顶级策略，即威胁行为者使用的一种广泛方法. MITRE 公司基于现实中发生的真实攻击事件，创建了一个对抗战术和技术知识库 ATT&CK(adversarial

Table 2 Comparison of Common Attack Threats and Defense Adversarial Models

表 2 常见攻击威胁与防御对抗模型对比

攻防视角	模型	阶段数	步骤
攻击者视角	网络杀伤[42]	7	侦察→武器化→交付→利用→安装→命令与控制→行动
	在线操作杀伤链[43]	10	获取资产→伪装资产→收集信息→协调与计划→测试防御→逃避检测→无差别接触→针对性接触→渗透资产→长期驻留
	MITRE ATT&CK[44]	14	侦察→资源开发→初始访问→执行→持久化→权限提升→防御绕过→凭证访问→发现→横向移动→收集→命令与控制→数据窃取→危害
防御者视角	IDDIL/ ATC[46]	7	发现阶段：识别资产→定义攻击面→分解系统→识别攻击向量→列出威胁源和攻击代理； 实施阶段：分析与评估→分类→控制
	MITRE ENGAEG[47]	9	规划→收集→检测→防御→转移→破坏→保证→激励→分析
	网络空间欺骗链[48]	8	制定目标→收集网络信息→设计封面故事→计划→准备→执行→监控→加固
	NIST 网络安全框架[45]	5	识别→保护→检测→响应→恢复

tactics, techniques, and common knowledge)[44]. ATT&CK 框架包括了攻击者在攻击过程中使用的 190 多项技术、400 多项子技术，其一经推出便得到了业内的广泛关注.

2.2 防御者视角

从防御者的角度来看，防御的目的是为了减轻网络攻击带来的损失，利用现有防御资源最大化提升系统的安全级别. 欺骗防御技术加强了在对抗攻击者时的主动性，通过创建虚假或具有误导性的信息，能够诱导和捕获攻击者，甚至通过蜜标等追踪型欺骗手段，进而实现对攻击者的溯源和反制.

2013 年，在美国发布的“改善关键基础设施网络安全”行政命令的指导下，NIST 网络安全框架[45]被提出，并在 2014 年的《网络安全增强法案》中得到了进一步的扩展. NIST 网络安全框架的核心结构包括 5 个核心结构，分别为识别、保护、检测、响应和恢复. IDDIL/ATC[46]是一种基于威胁驱动的网络安全分析模型，如图 3 所示，该方法分为 2 个工作阶段，其中 IDDIL 为发现阶段，ATC 为实施阶段. IDDIL/ATC 模型可作为 Cyber Kill Chain 模型的补充，为防御者提供了应对未知威胁的处理方法.

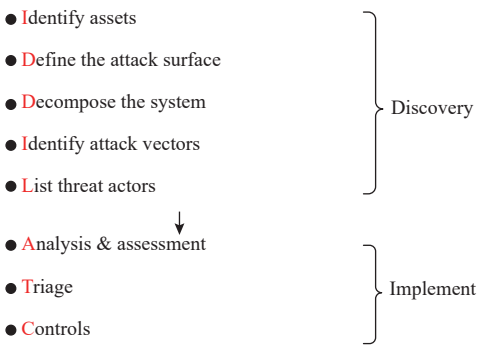


Fig. 3 Core concepts of IDDIL/ATC[46]

图 3 IDDIL/ATC 核心概念[46]

MITRE Engage[47]模型在 MITRE Shield 框架的基础上进行精简，将目标集中于对网络攻击的拒绝和欺骗 2 方面. 与 MITRE ATT&CK 框架类似，也是以矩阵的方式对防御战术和技术进行展示. Engage 矩阵从战略活动与作战活动 2 方面展开，并将作战活动具体划分为作战目标、作战方法和作战技术. 矩阵中的相关技术能够与 ATT&CK 的技术进行映射，更加清晰地展示防御者针对具体攻击手段可以采取哪些欺骗性手段. 图 4 展示了多阶段网络攻击中攻防技术的示意图.

3 基于攻击图的欺骗防御

攻击图通过将网络拓扑与漏洞建立起关联关系，对网络中潜在的攻击路径与脆弱点进行识别和风险评估. 攻击图作为一种网络安全建模工具，能够帮助防御者建立起对当前系统的一个全局性的视图，让防御者能够从理性攻击者的角度去分析最具威胁的攻击路径及行动序列. 为了提升欺骗防御的欺骗效果，做到有的放矢，加强对攻击威胁的感知和探查能力，同时把有限的资源部署在关键的脆弱点上，攻击图与欺骗防御的结合是目前的研究热点，也是未来欺骗策略优化的主要方向之一[49-53]. 如图 5 所示，本节将介绍欺骗防御中涉及的各种攻击图模型.

3.1 状态攻击图

状态攻击图中的节点通常表示网络状态信息，如系统的安全状态、攻击者的访问级别以及系统中的特定漏洞或配置. 边代表从一个状态到另一个状态的转移，通常对应于攻击者的行动[54-56]. 然而随着网络规模的增加，状态攻击图可能会变得非常复杂，该图包含数以千计的节点和边，将导致攻击图变得难以理解，因此目前已少有针对状态攻击图的研究.

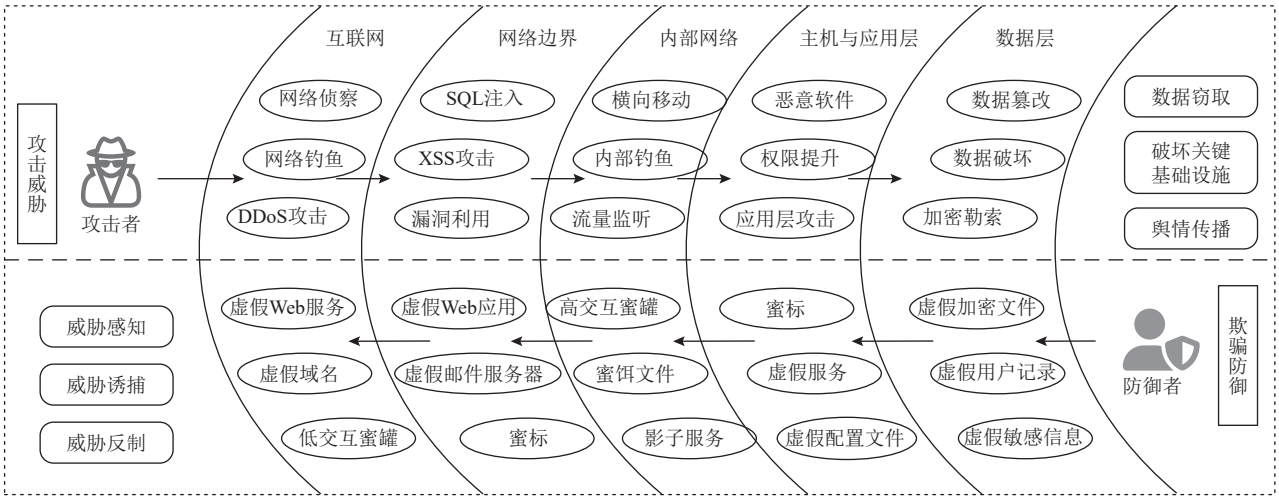


Fig. 4 Schematic diagram of offensive and defensive techniques in multi-stage network attacks

图 4 多阶段网络攻击中攻防技术示意图

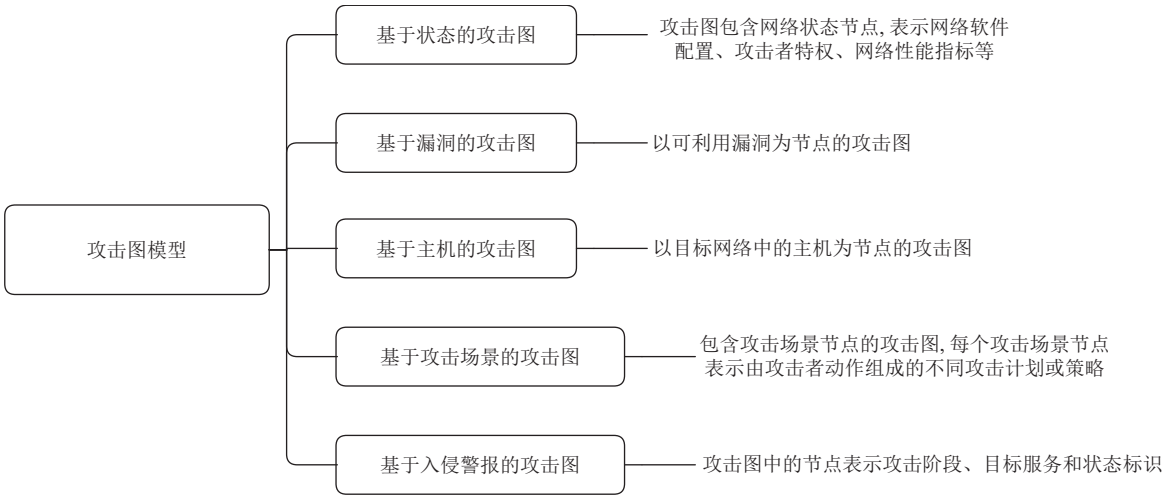


Fig. 5 Introduction of common attack graph models

图 5 常见攻击图模型介绍

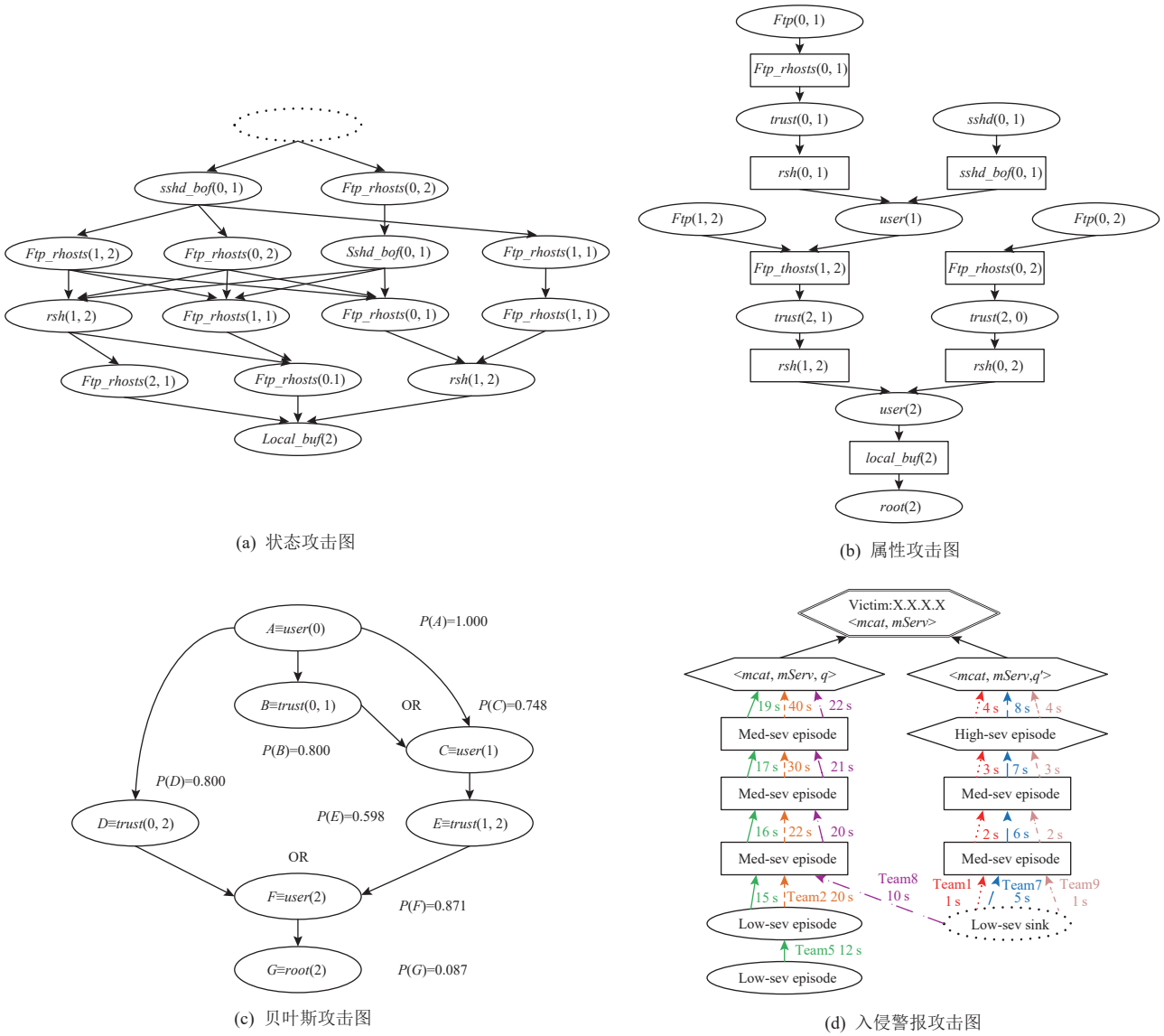
3.2 属性攻击图

属性攻击图相比于状态攻击图，其生成速度快，且结构相对简单，改善了状态攻击图中存在的状态爆炸问题. 如图 6 所示，属性攻击图一般包含 2 类节点：一类为条件节点，表示攻击者当前拥有的权限；另一类为漏洞节点，表示存在漏洞的服务以及利用漏洞后攻击者能够获得的相应权限. 由于属性攻击图相比于状态攻击图具备多种优势，目前的研究工作重点集中于属性攻击图[57-59].

3.3 概率攻击图

概率攻击图在传统攻击图的节点与边上增加概率值来量化攻击成功的可能性，由此计算攻击路径发生的概率和节点被攻陷的概率. 概率攻击图中的边则表示了节点间的因果关系，根据初始节点的概率值和节点间的因果关系可推导出后续所有节点的条件概率. 而概率的赋值通常基于历史数据、攻击复杂性、漏洞利用难度等. 通过该种量化方式，能够帮助防御者更加清晰地理解网络中众多可能攻击路径的风险大小，明确防御部署的重点方向.

Wang 等人[60]针对企业网络中存在多个可利用漏洞的复杂网络攻击问题，使用 MulVAL[61]工具生成攻击图，利用 CVSS 评分体系为攻击图中的每个节点引入全新的指标，计算攻击成功的条件概率. 通过概率攻击图展示了如何使用定量的指标来明确修补漏洞的优先级. 针对内部攻击者的攻击行为具有多步骤与伪装性强的特点，且攻击图模型对于单步检测结果存在不确定性，陈小军等人[62]在攻击图模型中引入转移概率表，根据已有的静态漏洞分析知识库和安全监控系统动态生成的事件置信度来确定条件概率表，设计了攻击意图推断算法，从而确定具有最



(a) 状态攻击图　(b) 属性攻击图

(c) 贝叶斯攻击图　(d) 入侵警报攻击图

Fig. 6　Common attack diagram[51-53]

图 6　常见攻击图[51-53]

大概率的攻击路径. 实验结果通过对内部攻击者的潜在的攻击意图进行推断, 能够减少不可信报警数量, 增强网络管理员对内网环境的感知能力.

3.4　贝叶斯攻击图

贝叶斯攻击图是一种特殊类型的概率攻击图, 它使用贝叶斯网络来表示攻击路径和状态之间的概率关系. 贝叶斯攻击图中着重强调了先验知识和条件概率的使用. 贝叶斯攻击图具有 3 个特点: 1)结合先验知识和新的观察数据来更新攻击概率; 2)使用条件概率来表示攻击事件和状态的依赖关系; 3)基于观测到的结果动态地更新攻击概率[63-66]. 图 7 总结了攻击图在网络安全中的用途, 其中贝叶斯攻击图常用于网络安全加固和欺骗防御资源的部署.

4　基于博弈论的欺骗防御

博弈论目前已广泛应用于网络安全问题的建模, 尤其是针对攻击者与防御者间的持续性对抗. 博弈论方法在假设攻防双方理性的前提下, 通过寻找纳什均衡, 为参与者推荐最大化自身收益的策略. 如图 8 所示, 展示了欺骗防御中攻防博弈建模的核心概念, 本节将对相关研究工作进行总结和介绍.

4.1　参与者

目前大多数针对网络攻防进行建模的研究工作基本将博弈的参与方分为攻击者和防御者. 攻击者采用网络侦察、社会工程攻击、恶意软件部署、权限提升和横向移动等一系列攻击手段对目标系统进行

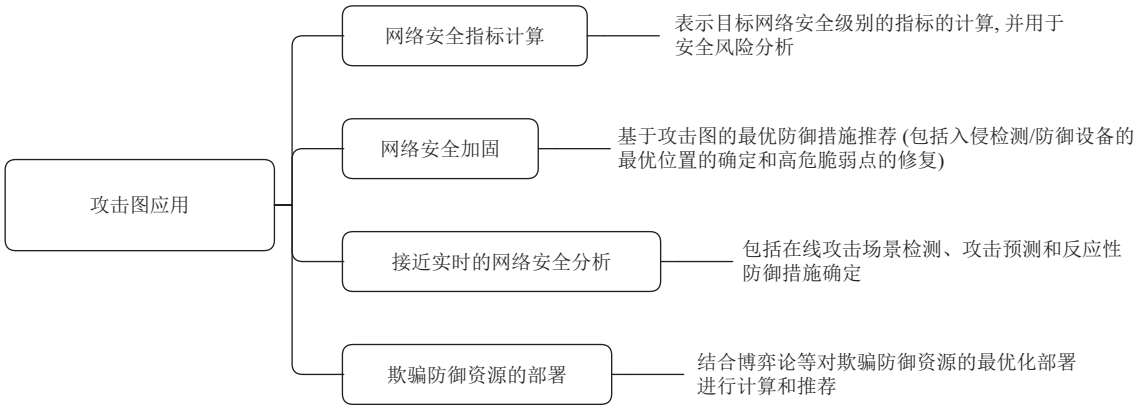


Fig. 7 Summary of the purpose of the attack graph

图 7 攻击图的用途总结

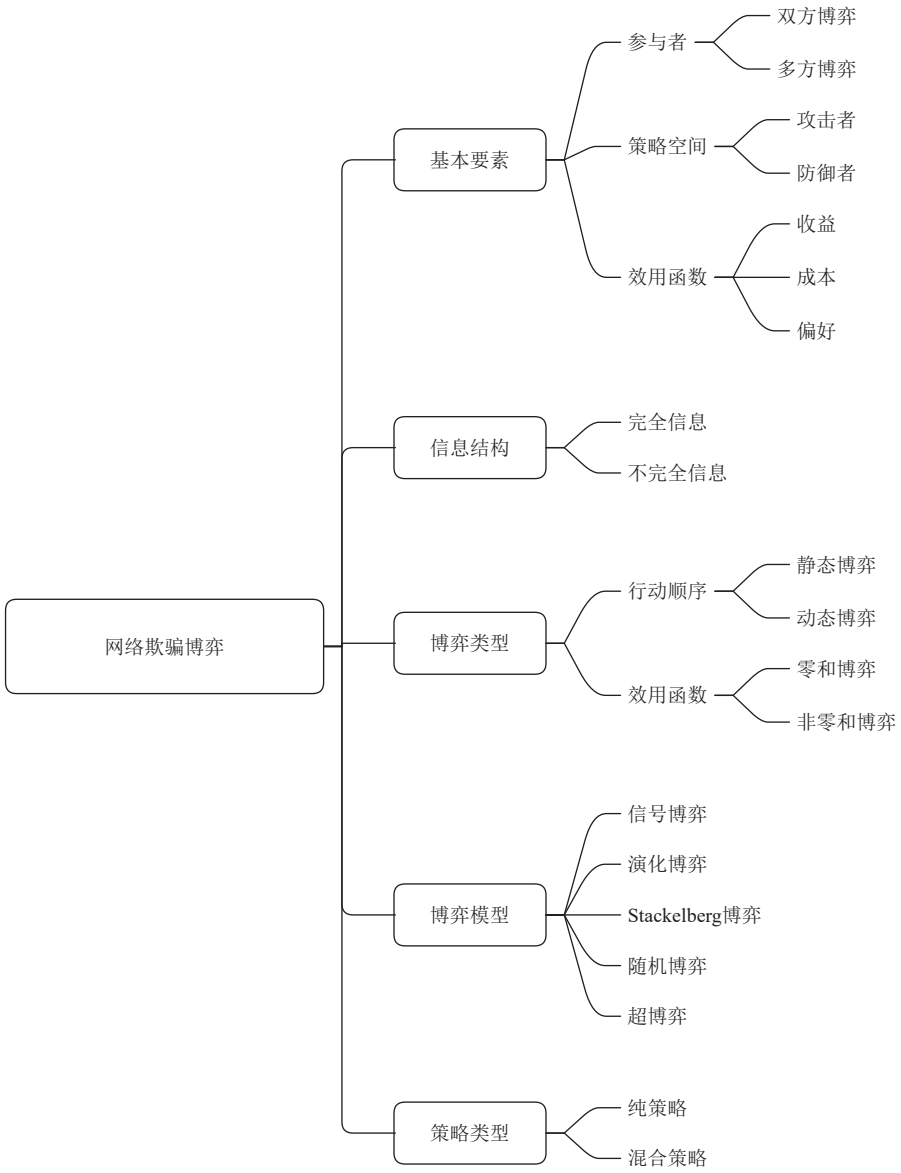


Fig. 8 Classification of network deception game model

图 8 网络欺骗博弈模型分类

破坏与监控. 防御者则通过放置蜜罐、添加虚假信息、变幻网络拓扑等欺骗性动作误导或迷惑攻击者.

Anwar 等人[67] 建立了一个双人零和博弈模型, 用于指导防御者如何对蜜罐和蜜饵进行分配, 并根据网络拓扑结构以及节点的重要性对蜜罐放置的策略进行了优化. 而在文献 [68] 中, 研究人员将网络欺骗视为一个独立于防御者和攻击者之外的博弈参与者, 构建了一个三方进化博弈模型. 此时网络欺骗系统通过诱导攻击者攻击虚假漏洞, 进而捕捉和分析攻击者的进攻意图, 并将结果传递给防御者, 帮助防御者发挥自身优势策略, 减少防御成本开销和攻击者的收益. 仿真结果表明, 网络欺骗系统可以在攻防博弈的过程中, 利用捕获的攻击者信息有效地改进入侵检测系统对攻击的检测概率. 为抵抗 DDoS 攻击, Zhou 等人[69] 采用基于 shuffling 的移动目标防御方法, 通过在攻击者、防御者和用户之间构建三方博弈来寻找最优的 shuffling 策略, 利用多目标马尔可夫决策过程解决了 shuffling 的有效性和成本之间的权衡问题.

4.2 策略空间

在真实世界的网络攻击中, 攻击者具备的战术与技术复杂多样, MITRE ATT&CK 知识库中涵盖了 APT 攻击者的绝大多数 TTPs, 同样地与之对应的 MITRE Engage 框架专注于网络欺骗与网络阻断技术, 从对抗前的准备、对抗过程中的欺骗与阻断, 到对抗后的回顾分析和威胁情报收集, 为防御者提供了众多有效的防御策略. 但在将网络攻防对抗建模为博弈模型时, 为了更好地进行理论分析, 通常会将这种复杂性进行简化, 以便于更清晰地阐述攻防策略的动态变化和可能导致的结果. 另外, 随着攻防动作的增多, 模型的复杂度也随之上升, 极大地增加了寻找博弈模型纳什均衡解的难度, 可能同时出现多个局部均衡, 甚至受到计算复杂度的限制导致无法求解均衡. 因此, 为了缩小策略空间, 便于更高的维度上理解攻防交互的本质, 目前多数研究会围绕某一类型的威胁展开博弈建模.

Thakoor 等人[70] 针对网络侦察与扫描攻击, 利用 Stackelberg 博弈对攻防双方进行建模, 防御者通过欺骗策略掩盖网络中主机的真实配置, 攻击者则根据观察结果和预期收益对网络中的主机选择性攻击. 文献 [71] 提出了一种 APT 欺骗博弈模型来描述防御者通过在内网部署诱饵资源来对抗攻击者在内网中的横向移动. 由于目前的防御策略基本上是基于已知的攻击方式制定, 面对未知的零日漏洞攻击的挑战, 通常难以进行有效的防御. 而欺骗防御为对抗此类威胁提供了一种可行的方法, 攻击者一旦被欺骗手段误导, 将零日漏洞等高级攻击手段暴露在诱饵环境中, 防御者便可以在真实的攻击发生前捕获攻击行为, 为加固系统争取宝贵的时间. Sayed 等人[72] 将蜜罐的分配方法作为欺骗策略, 设计了一种双人零和博弈模型来研究攻击者可能使用的潜在侦察和攻击路径. 考虑到零日漏洞的存在, 攻击者可能通过创建未知的攻击路径来躲避蜜罐的诱捕, 因此引入了敏感性分析来讨论不同的零日漏洞对欺骗技术性能的影响, 并在此基础上提出了多种针对零日攻击威胁的缓解策略.

4.3 效用函数

在攻防对抗中, 影响参与者决策的关键要素是行动带来的预期收益, 只有充分地了解自身和对手的效用函数, 才能做出最大化自身收益的决策. 攻击者实施成功的网络攻击对系统造成破坏、对数据进行窃取等行为将为其带来正收益, 而攻击被 IDS 等检测、行为被蜜罐捕获而引起告警, 则收益视为负. 在零和博弈模型中, 攻防间的收益是相对的, 一方的收获即为另一方的损失. 在计算收益时, 攻防成本也常常被考虑在效用函数中. 攻击者的成本通常包括利用漏洞、开发恶意软件、执行攻击时的资源消耗等. 而防御者在检测攻击时部署的检测设备、执行欺骗策略时放置的蜜罐数量等, 均需要消耗大量的防御资源[73]. Sengupta 等人[74-75] 使用通用漏洞评分体系 (CVSS) 对博弈中的每个状态的平均效用进行量化计算, 构建了一个一般和马尔可夫博弈且用于解决在云网络中如何利用移动目标防御方法有效检测多阶段攻击的问题.

4.4 完全信息和不完全信息

在完全信息的假设下, 攻防双方都完全了解对方的能力、策略和意图, 但该假设在现实世界的网络攻防中难以实现, 防御者不可能完全了解所有潜在的威胁和漏洞, 而攻击者对于部署了欺骗资源的目标系统也无法做到完全掌握. 因此, 在对网络欺骗博弈进行建模时, 大多数研究工作都将不完全信息作为前提假设.

Huang 和 Zhu[76] 从完全信息到不完全信息、从静态博弈到多阶段动态博弈, 以及从对称到非对称信息结构等不同角度, 对网络欺骗博弈进行分析, 将攻击者类型与欺骗技术结合, 提出了一种对抗 APT 的欺骗防御理论框架. 杨峻楠等人[77] 针对物联网安全中的欺骗防御策略进行了研究, 并提出在防御者利用蜜罐作为欺骗工具来诱骗攻击者的同时, 攻击者

也通过不同类型的攻击来迷惑防御者，即通过“佯攻”的方式扰乱防御者的观测结果. 因此将此类双方均存在欺骗行为的问题建模为不完全信息的贝叶斯博弈，研究结果表明，攻击者的主动进攻频率存在一个阈值，一旦高于该阈值，攻防双方均将采用欺骗策略.

4.5 零和博弈与非零和博弈

在零和博弈中，通常假设攻击者和防御者间的利益是完全对立的，即一方的损失完全等于另一方的收益. 文献 [78] 提出了零和单边部分可观察随机博弈(OS-POSG)模型，其中防御者战略性地将蜜罐放置在物联网网络中，以欺骗攻击者的行动并减轻僵尸网络的传播. 而非零和博弈则会更加反映了现实世界网络攻击的复杂性，当考虑到攻击和防御的成本时，双方的效用函数结构通常是不同的. Thakoor 等人[79] 提出一种非零和的网络伪装博弈模型，为了解决均衡求解困难的问题，设计了一个基于 MILP 的算法以及多种加速计算的技术，以此给出了一个完全多项式时间的近似解方案.

4.6 静态博弈和动态博弈

静态博弈是指攻击者与防御者几乎同时做出决策，在欺骗防御的相关研究中，攻击者根据当前侦察的结果对目标系统做出攻击与不攻击的决策，而防御者则预先设定并部署一定数量的欺骗资源，双方的决策均不受历史行为的影响. 由于主动欺骗防御更倾向于采取多阶段手段对威胁进行诱捕，且高级威胁者往往会根据观测的结果动态调整攻击策略，因此动态博弈虽然增加了模型的复杂性，但更贴近实际的欺骗防御场景，在文献 [80−82] 中通过采用动态欺骗博弈方法对网络侦察、DDoS 攻击和工业物联网中的 APT 攻击进行了分析与建模.

博弈论与攻击图的结合帮助防御者可以根据攻防态势的变化对欺骗防御的资源部署做出最优化的决策，在最优资源成本规划下，实现较为理想的安全防护，表 3 对基于攻击图与博弈论的欺骗防御代表性工作进行了总结与归纳.

5 基于大模型的欺骗防御

5.1 AIGC 在网络安全领域的应用

AIGC[94](artificial intelligence generated content)指的是通过人工智能技术自动生成包括文本、图像、音频和视频等多种形式的媒体内容. 大语言模型[95] 作为实现 AIGC 的一种方式，其通过分析和学习大量数据来生成文本. 近年来，随着大语言模型技术的持续性火热，学术界和工业界都对其赋能，并在网络安全领域进行了一系列探索. 如表 4 所示，列出了 3 款较为热门的基于 ChatGPT 开发的产品的功能特点.

微软的 Security Copilot 利用 GPT-4 总结安全事件、识别漏洞和受影响的账户[99]. SecurityLLM[100] 是一个专门为网络安全威胁检测设计的预训练语言模型，其包含 2 个关键部分：SecurityBERT 和 FalconLLM. SecurityBERT 用于网络威胁检测，而 FalconLLM 用于事故响应和恢复系统. 实验分析显示，该方法能够以 98% 的总体准确率检测到 14 种不同类型的攻击.

VWC-MAP[101] 框架利用大模型技术，通过 2 层分类方法实现自动识别功能：第 1 层将漏洞分类到弱点，第 2 层将弱点分类到攻击技术，实现了自动准确地识别出与特定漏洞相关的所有攻击技术. 实验结

Table 3 Summary of Deception Defense Work Based on Attack Graph and Game Theory

表 3 基于攻击图与博弈论的欺骗防御工作总结

文献 年份	攻击图类型	攻击类型	应用目标	欺骗技术	博弈模型
[83] 2020	多层攻击图	漏洞利用	网络安全加固、最优防御策略、最小化防御成本、预测攻击路径	MTD	
[84] 2020	贝叶斯攻击图	内部威胁	最优防御策略	MTD	动态三方博弈
[85] 2020	漏洞依赖图	侦察	最优欺骗策略	网络欺骗	POSG、超博弈
[86] 2020	有向无环图		最优欺骗策略	添加诱饵资源	Stackelberg 博弈
[87] 2020	贝叶斯攻击图	漏洞利用	最优欺骗策略	蜜罐、诱饵节点	
[88] 2021	多层攻击图	侦察，漏洞利用	最优欺骗策略	蜜罐	信号博弈
[89] 2022	概率攻击图	漏洞扫描、漏洞利用	最优安全加固成本		
[90] 2022	概率攻击图	APT	网络安全风险评估、最优安全资源分配		MDP
[91] 2022	Active Directory 攻击图	Active Directory 攻击	网络安全加固、有限预算下的最优防御策略		Stackelberg 博弈
[92] 2023	概率攻击图	漏洞利用	最优诱饵资源分配	蜜罐、蜜饵	MDP、非零和博弈
[93] 2023	概率攻击图	侦察	网络安全加固	MTD	Stackelberg 博弈、MDP

Table 4 Products Developed Based on ChatGPT

表 4 基于 ChatGPT 开发的产品

产品	功能
ChatPDF[96]	一个用于帮助理解文档（PDF）内容的网页应用，只需将 PDF 文件上传到 ChatPDF，聊天机器人将会自动提供一个摘要，并建议提出问题，以了解更多关于该文件的信息.
Auto-GPT[97]	一种基于 ChatGPT API 的 AI 代理，它能够自动执行由自然语言描述的任务. 通过将目标拆解为子任务并自动进行网络搜索和数据收集，利用 GPT 进行文件存储与总结，最终实现任务的自动化执行.
PentestGPT[98]	一款由 ChatGPT 赋能主要针对 web 渗透测试的工具. 旨在自动化渗透测试过程中，以交互方式运行指导渗透测试人员进行具体操作.

果经过网络安全专家的交叉验证，显示 VWC-MAP 能够将漏洞与弱点类型关联起来，准确率高达 87%，并且能够将弱点与新的攻击模式关联起来，准确率高达 80%.

虽然目前将网络安全与大模型结合的工作并不少见，但将其应用于欺骗防御领域的研究目前仍处于探索阶段，相关成果较少. 本文针对传统 Web 蜜罐的局限性，利用大语言模型的优势，提出了一种基于大模型的智能化外网蜜点生成技术.

5.2 基于大模型的外网蜜点生成技术

5.2.1 蜜点基本概念

蜜点[102]是一种区别于传统蜜罐的新型欺骗防御技术，其摒弃了主动吸引攻击者前来攻击的思路，主要针对 APT 攻击者，依据“以未知应对未知，以隐蔽应对隐蔽，以威胁应对威胁”的核心思想进行设计开发. 蜜点的具体优势有 3 个方面：

1）隐蔽性. 蜜罐采用主动式诱骗，通过暴露明显的漏洞引诱攻击者探查、观察攻击行为，其缺点是难以应对高级攻击者，且易被反利用；蜜点则是被动式诱捕，强调未知性和隐蔽性，其通常部署于被保护目标周围，伪装成正常主机或服务，进而增强对隐蔽攻击者的捕获几率.

2）多样性. 蜜点的存在形式更为丰富，从外部网络到内部网络、从主机层到数据层，可依据不同的保护目标，针对性地部署不同类型的蜜点.

3）动态性. 传统的蜜罐通常采用静态化部署方式，容易被谨慎的攻击者识别并成为攻击者发动攻击的跳板；蜜点可通过智能决策算法，从部署数量、位置、仿真类型等多个维度动态调整，从而扰乱攻击者的侦察结果.

定义 2. 外网蜜点. 一种专为外部网络环境而设计，通过模拟企业在外部网络上公开的真实业务，并配合具有一定迷惑性但不会对合法用户产生混淆的蜜点域名的轻量级欺骗防御装置.

5.2.2 传统 Web 蜜罐的局限性

目前，市场上流行的 Web 蜜罐捕获攻击的效果常常与其仿真度成正比，通常存在 3 个问题：

1）缺乏灵活性. 大部分 Web 蜜罐是基于主流 CMS 前端模板制作而成，针对被保护企业的重点业务进行定制化处理. 虽然保证了仿真度，但缺乏变化的能力.

2）人工成本高. 定制化的蜜罐需要较高的人工成本进行模拟仿真，针对不同的业务类型，需定制不同类型服务的蜜罐产品，极大地增加了人工参与的程度.

3）即时性不足. 随着网络环境的持续动态变化和业务需求的调整，需要满足用户的即时性需求，传统的蜜罐产品无法满足快速适应新需求和即插即用的场景.

结合上述传统 Web 蜜罐的不足，以及完全依赖传统爬虫获取网站资源模拟 Web 服务，会产生大量报错、需要花费大量的时间用于人工调试、效率较低等问题.

本文结合了大模型技术，充分利用其文本处理能力强，代码解析全面，结合少量人工输入，便能动态、自适应地应对不同的场景需求等优势，结合蜜点技术，提出了一种基于 ChatGPT 的外网蜜点生成算法，蜜点的生成流程如图 9 所示. 具体的算法如算法 1 所示.

算法 1. 外网蜜点页面生成算法.

输入：目标网站主要视觉元素集 S(图片、颜色、页面占比、页面样式、文字内容)，与大模型对话的初始 Prompt 模板 P，预期实现功能 I；

输出：具备功能 I 的蜜点页面.

① $Backendcode \overset{GPT}{\leftarrow} P \vee I$；

/* 将预期实现功能 I 与初始 Prompt 模板 P 结合输入到 GPT 中，生成后端逻辑代码*/

② for s in S

/* 将目标网站的主要视觉元素依次从集合 S 中进行提取*/

③ $NewBackendcode \overset{GPT}{\leftarrow} Backendcode \vee s$; /* 在对话长度受限的情况下，依次分析页面中某个元素*/

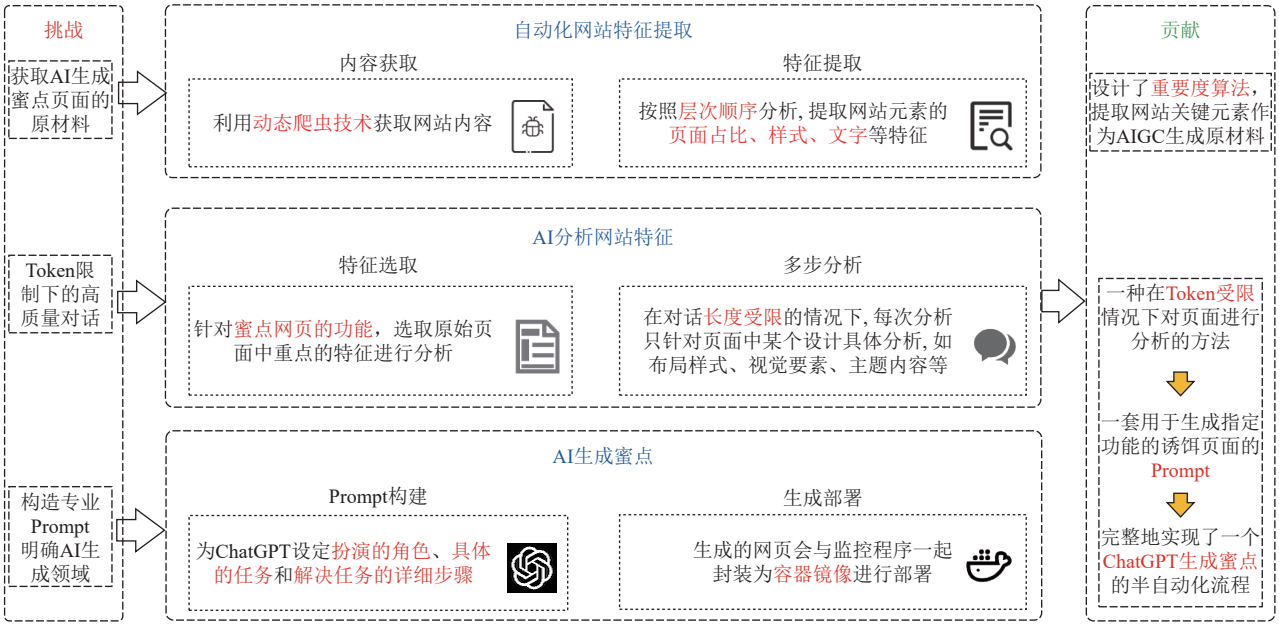


Fig. 9 AIGC-assisted HoneyPoint generation workflow

图 9 AIGC 辅助蜜点生成的工作流程

④ *Backendcode* ← *NewBackendcode*;

⑤ end for

⑥ *index.html* ← *Backendcode*./*将蜜点页面以 html 格式保存在本地 */

图 10 展示了关于生成蜜点诱饵页面的 prompt 模板片段. 在 role_play_prompt 中，GPT 被引导扮演一位经验丰富的前端工程师，负责将客户需求转化为吸引人且用户友好的 HTML 页面. 该部分强调了在不需要用户额外输入的情况下，创造符合客户规格要求的解决方案的重要性. 在 task_prompt 中，GPT 具体的任务是创建一个包含特定功能(page_func)的网页，该功能主要反映出某个给定标题(title)页面的视觉元素和内容主题. 目标是保持网站内部页面间的连贯性，让用户认为其在同一个网站内进行浏览. 生成一个网页的具体步骤为：

1)分析网页图片信息将其放在正确的位置.

2)分析网页的主要视觉颜色，确保诱饵网页的主视觉颜色与其相同.

3)基于分析的结果生成完整的网页代码.

通过动态爬虫采集对应页面的关键元素，利用算法 1 生成诱饵页面，并将其与准备好的监控程序打包成 docker 镜像，给后续部署的外网蜜点使用. 通

```
role_play_prompt =" " "
You are an accomplished frontend engineer, deftly translating client reguirements\
into visually compelling, highly functional, and user-friendly HTML pages.\
Your solutions should always be innovative and efficient, without requiring \
additional input from the use. Leverage your strengths as an accomplished \r
frontend engineer to deliver elegant, user-friendly HTML pages that meet client specifications.
" " "
task_prompt=f" " "
Your task is to create a page with a"{page_func}" feature, mirroring the visual elements and content theme \
of a page titled"{title}". The goal is to ensure continuity, making users feel they ' re within the same website.
You can achieve this by following these steps:
1. Analyze the image information of the page and consider how these images can be used on the imitative page.
2. Analyze the color information of the page and determine the colors to be used on the imitative page.
3. Based on the analysis results, write code that meets the reqUirements.
" " "
system_prompt =role_play_prompt+task_prompt
```

Fig. 10 HoneyPoint decoy Web pages generated by prompts code snippet

图 10 生成蜜点诱饵网页的 prompt 代码片段

过算法 1 的实现，将 docker 镜像与蜜点封装技术进行了融合，形成可供用户自主选择生成特定对象的诱饵网页的外网蜜点，例如用户希望生成目标网站为“https://newcas.gzhu.edu.cn/”的外网蜜点，利用本文提供的框架，其诱饵页面的仿真效果如图 11 所示，仿真好的蜜点页面可以进一步封装为蜜点镜像供后续使用.

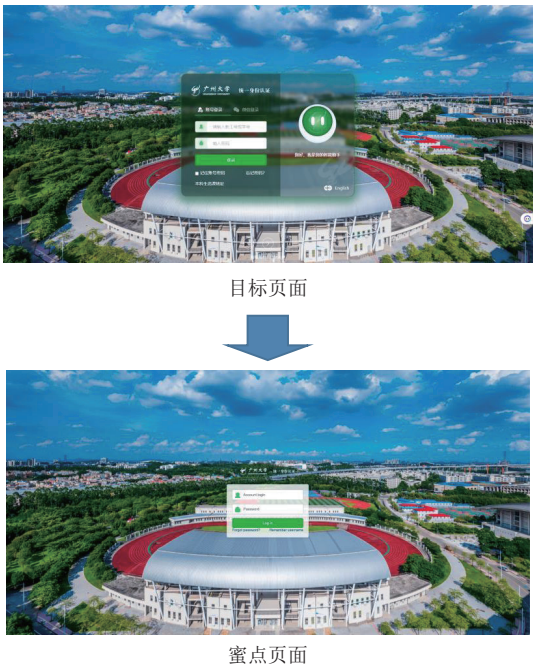


Fig. 11 AIGC assisted in generating HoneyPoint page effect display

图 11 AIGC 辅助生成蜜点页面效果展示

5.2.3 实验结果与分析

为验证本文提出的基于 ChatGPT 的外网蜜点生成技术的有效性，生成 8 个外网蜜点并将其部署在云服务器上，将一个 Snare 蜜罐也部署在公网，进行数据收集与结果比较，网络拓扑图如图 12 所示.

1)页面仿真性对比

通过对同一目标网站进行模拟，以广交会主站页面为例，效果如图 13 所示. 可以看出本文方法生成的外网蜜点页面在稳定性上有一定优势，通过特别

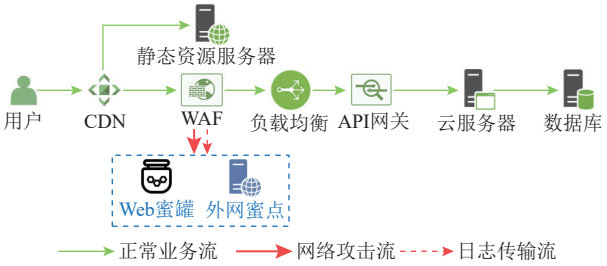


Fig. 12 Experimental network topology diagram

图 12 实验网络拓扑图

(a) 传统蜜罐模拟Web服务报错

(b) 基于AIGC生成的蜜点页面

Fig. 13 Comparison between our method and the traditional Honeypot simulation Web service

图 13 本文方法与传统蜜罐模拟 Web 服务的对比

设置的提示词模板，所仿真的蜜点页面不会与原始页面保持相同，而是具有独有的特色. 与 Snare 蜜罐的具体对比结果如表 5 所示.

2)有效性分析

对部署的 8 个外网蜜点捕获的攻击 IP 进行分析，收集 24 h 的数据，其中外网蜜点共捕获 303 个 IP，踩

Table 5 Comparison of Our Method with the Generation Quality of Snare Honeypots

表 5 本文方法与 Snare 蜜罐的生成质量对比

比较条目	Snare 蜜罐	本文方法
页面生成	采用传统爬虫的方式实现 Web 服务模拟.	在传统爬虫技术的基础上结合大语言模型.
仿真程度	仿真程度低，爬取时常因样式文件依赖关系导致页面无法显示.	仿真程度较高，通过精心设计的提示词，利用大语言模型深度分析.
稳定性	样式文件爬取缺失时出现前端显示不全，甚至产生报错.	利用大模型的特征分析、语言理解和代码构建能力，生成具有独立特色的模拟页面.
灵活性	页面生成后不易更改.	可根据用户需求定制化更新模拟页面.

蜜次数为 2 126 次. 对整体踩蜜 IP 进行 AbuseIPDB 查询, IP 的威胁值分布如图 14 所示. 威胁值接近 100 的踩蜜 IP 占总 IP 的 29.7%, 且踩蜜 IP 中按地理位置分析, 如图 15 所示, 来源于中国和美国的 IP 数为前 2 位, 其中来源于中国的踩蜜 IP 威胁值较低的占比居多, 而来自 IP 地理位置标记为美国的, 其威胁值多数为接近 100.

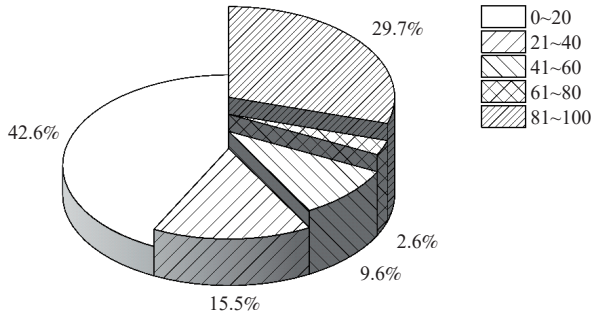


Fig. 14 IP threat value distribution captured by the HoneyPoint of the external network

图 14 外网蜜点捕获的 IP 威胁值分布

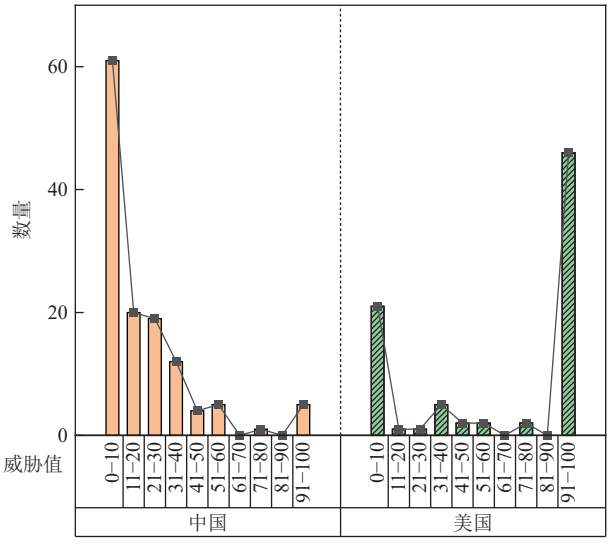


Fig. 15 Comparison of threat value distribution between Chinese and American IP addresses

图 15 中美 IP 地址威胁值分布对比

通过对同时段 Snare 蜜罐捕获的攻击 IP 进行分析, 24 h 内共捕获 29 个攻击 IP, 探查蜜罐 80 次, 且其中绝大多数为扫描行为. 相比较, 外网蜜点平均单个蜜点踩蜜次数接近 266, 平均捕获 IP 数为 38 个, 均高于传统蜜罐. 由此可见外网蜜点在捕获平时不易被发现的网络扫描等攻击行为时具有一定的优势.

通过对同时段 Snare 蜜罐捕获的攻击 IP 进行分析, 24 h 内共捕获 29 个踩蜜 IP, 探查蜜罐 80 次, 且其中绝大多数为扫描行为. 相比较, 外网蜜点平均单个蜜点踩蜜次数接近 266, 平均捕获 IP 数为 38 个, 均高于传统蜜罐. 由此可见外网蜜点在捕获平时不易被发现的网络扫描等攻击行为时具有一定的优势.

3)踩蜜 IP 攻击行为分析

在外网蜜点捕获的 IP 中, 存在多个具有明显恶意行为的攻击. 其中“65.xx.102.xx”远程下载恶意代码并执行, 疑似墨子僵尸网络利用 GPON 漏洞传播, 蜜点记录日志如图 16 所示. 多个踩蜜 IP 存在利用弱口令进行网关服务扫描行为, 使用的弱口令为 admin/Feefifofum.

Fig. 16 Attack IP log recorded by HoneyPoints

图 16 蜜点记录的攻击 IP 日志

5.2.4 欺骗防御技术展望

通过对近年来基于攻击图模型与博弈理论在欺骗防御中研究工作的回顾, 以及大模型驱动下为新型欺骗防御技术的发展带来了新的机遇与挑战. 如何使欺骗防御决策更加合理化、智能化成为了当前以及未来需要重点考虑的问题. 在基于蜜点的欺骗防御技术的未来工作方面, 我们从蜜点的隐蔽性、多样性与动态性 3 个方面提出 3 点思考和展望:

1)结合大模型增强蜜点的隐蔽性. 蜜点的优势在于其具有较强的隐蔽性, 网络蜜点能够根据周边环境的服务信息生成与之相近的蜜点服务. 在大模型的辅助下, 用户可自主对蜜点进行定制化配置, 同时蜜点仿真程度也将得到增强. 通过大模型与蜜点技术的结合, 蜜点服务的生成效率及适配性将进一步提升, 攻击者更加难以区分蜜点服务与真实服务间的差别, 从而增大隐蔽攻击者探测踩中蜜点的几率, 以及形成对谨慎攻击者的威慑, 使其不敢轻易采取大范围的扫描侦察等行为.

2)提升欺骗诱捕的甜度. 除了本文工作中提出的外网蜜点, 蜜点的存在形式具有多种类型. 根据适应环境的不同, 需要使用多种欺骗技术对威胁进行诱捕. 通过将大模型与网络安全领域相结合, 从欺骗防御角度进行针对性训练, 能够制作和生成甜度更高的蜜饵, 如邮件蜜点中根据特定主题自动化生成高度契合的邮件内容, 可增大对攻击者的迷惑性; 账

号蜜点中依据种子用户名，可利用大模型生成一系列高仿真的域用户账号，通过对蜜点账号进行异常行为监控，对网络中横向移动的攻击者进行提前预警与识别；文件蜜点中文件名称、文件内容以及文件放置的路径等均可利用大模型进行辅助指导与生成.

同时随着大模型技术的突破性进展，使人机交互迈向智能化发展的道路，从欺骗防御技术角度出发，利用大模型构建高交互的蜜点终端从而与攻击者进行深入交互成为了一种潜在的诱骗方式. 通过对大模型进行提示词语料库构建与提示词模板训练，使其能够生成对应攻击指令的响应结果，在蜜点终端命令行中与攻击者逐步交互，捕获更多攻击者的技术和战术，从而暴露潜在的攻击目的. 此外，也可将大模型应用于漏洞数据的分析，设计面向漏洞请求响应的大模型提示语，将其与蜜点技术结合，生成多种能够模拟漏洞利用响应的蜜点漏洞诱饵，针对漏洞利用攻击进行欺骗式响应，提升对攻击行为的粘随性.

3)欺骗防御策略的动态调优. 本文针对传统Web蜜罐存在的局限性，提出了一种利用大模型生成外网蜜点的技术. 然而蜜点生成后，面对与攻击者的交互，需要动态变化蜜点模拟的目标业务、蜜点所处的位置，以及需要增加或删除蜜点的数量. 否则，一成不变的蜜点随着时间的推移将会使攻击者有所察觉，失去其威胁探查的能力. 针对上述问题，本文提出蜜阵的概念.

定义 3. 蜜阵. 蜜阵是支持蜜点间协同联动的设备，其使用规范化的通信协议和统一的命名规则，针对安全态势的变化，输出最优化的蜜点部署和变化策略，实现蜜点间的互联互通，协同调度.

蜜阵的实现可通过攻击图模型作为导向，依据博弈理论在攻击者与防御者之间构建欺骗博弈模型，实时动态地做出最优的蜜点部署策略. 而大模型、人工智能、强化学习等理论与方法，可以为蜜阵协同调动蜜点的生成和变化提供有力的支撑. 我们将在后续的工作中进一步阐述蜜阵的模型构建与理论方法.

6 结束语

本文对欺骗防御领域的相关技术从威胁探查与智能决策的角度进行了系统性地回顾和总结. 面对高隐蔽未知威胁，从攻击图与博弈论2方面，讨论了如何最优化地运用欺骗防御技术，帮助防御者扭转攻防对抗中处于劣势的局面. 针对传统蜜罐难以应对当前复杂的网络威胁等不足，提出了一种基于大模型的智能化外网蜜点生成方法，通过实验验证了方法的有效性，和传统蜜罐相比也具有一定的优势. 最后通过对攻击图模型、博弈理论以及大模型方法的整体阐述，对融合上述技术的蜜阵技术实现进行了展望. 综上，形成了“以攻击图为依据，博弈理论为指导，欺骗防御为手段，大模型为辅助”的主动防御机制，增强了防御者威胁预测、威胁感知与威胁诱捕的能力，最终做到在攻守间运用欺骗技术，扭转乾坤、反客为主.

作者贡献声明： 王瑞负责文献收集与整理、数据分析、实验方案制定以及论文撰写；阳长江、邓向东负责完成实验部分；刘园负责论文的修订；田志宏负责整体架构规划并修改论文.

参 考 文 献

[1] Heckman K E, Stech F J, Schmoker B S, et al. Denial and deception in cyber defense[J]. Computer, 2015, 48(4): 36−44

[2] Wang C, Lu Zhuo. Cyber deception: Overview and the road ahead[J]. IEEE Security & Privacy, 2018, 16(2): 80−85

[3] Ren Yitong, Xiao Yanjun, Zhou Yinghai, et al. CSKG4APT: A cybersecurity knowledge graph for advanced persistent threat organization attribution[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(6): 5695−5709

[4] Zhou Yinghai, Ren Yitong, Yi Ming, et al. CDTier: A Chinese dataset of threat intelligence entity relationships[J]. IEEE Transactions on Sustainable Computing, 2023, 8(4): 627−638

[5] Butavicius M, Ronnie T, Simon J H. Why people keep falling for phishing scams: The effects of time pressure and deception cues on the detection of phishing emails[J]. Computers & Security, 2022, 123: 102937

[6] Stellios I, Kotzanikolaou P, Psarakis M. Advanced persistent threats and zero-day exploits in industrial Internet of things[G]//Security and Privacy Trends in the Industrial Internet of Things. Berlin: Spring, 2019: 47−68

[7] Horak K, Bosansky B, Tomasek P, et al. Optimizing honeypot strategies against dynamic lateral movement using partially observable stochastic games[J]. Computers & Security, 2019, 87: 101579

[8] Jiang Wei, Fang Binxing, Tian Zhihong, et al. Research on defense strategies selection based on attack-defense stochastic game model[J]. Journal of Computer Research and Development, 2010, 47(10): 1714−1723 (in Chinese)
(姜伟, 方滨兴, 田志宏, 等. 基于攻防随机博弈模型的防御策略选取研究[J]. 计算机研究与发展, 2010, 47(10): 1714−1723)

[9] Aydeger A, Manshaei M H, Rahman M A, et al. Strategic defense against stealthy link flooding attacks: A signaling game approach[J].

IEEE Transactions on Network Science and Engineering, 2021, 8(1): 751−764

[10] Han Xiao, Kheir N, Balzarotti D. Deception techniques in computer security: A research perspective[J]. ACM Computing Surveys 2018, 51(4): 1−36

[11] Baykara M, Resul D. A novel honeypot based security approach for real-time intrusion detection and prevention systems[J]. Journal of Information Security and Applications, 2018, 41: 103−116

[12] Sun Yanbin, Tian Zhihong, Li Mohan, et al. Honeypot identification in softwarized industrial cyber–physical systems[J]. IEEE Transactions on Industrial Informatics, 2020, 17(8): 5542−5551

[13] Franco J, Aris A, Canberk B, et al. A survey of honeypots and honeynets for Internet of things, Industrial Internet of things, and cyber-physical systems[J]. IEEE Communications Surveys & Tutorials, 2021, 23(4): 2351−2383

[14] Pawlick J, Zhu Quanyan. Game Theory for Cyber Deception[M]. Berlin: Springer, 2021

[15] Pawlick J, Edward C, Zhu Quanyan. A game-theoretic taxonomy and survey of defensive deception for cybersecurity and privacy[J]. ACM Computing Surveys, 2019, 52(4): 1−28

[16] Huang Yunhan, Zhu Quanyan. Deceptive reinforcement learning under adversarial manipulations on cost signals[C] //Proc of 10th Int Conf on Decision and Game Theory for Security. Berlin: Springer, 2019: 217−237

[17] Pourranjbar A, Kaddoum G, Ferdowsi A, et al. Reinforcement learning for deceiving reactive jammers in wireless networks[J]. IEEE Transactions on Communications, 2021, 69(6): 3682−3697

[18] Abolfathi M, Shomorony I, Vahid A, et al. A game-theoretically optimal defense paradigm against traffic analysis attacks using multipath routing and deception[C] //Proc of the 27th ACM on Symp on Access Control Models and Technologies. New York: ACM, 2022: 67−78

[19] Olowononi F O, Anwar A H, Rawat D B, et al. Deep learning for cyber deception in wireless networks[C] //Proc of Int Conf on Mobility, Sensing and Networking. Piscataway, NJ: IEEE, 2021: 551−558

[20] Gong Xueluan, Wang Qian, Chen Yanjiao, et al. Model extraction attacks and defenses on cloud-based machine learning models[J]. IEEE Communications Magazine, 2020, 58(12): 83−89

[21] Ferguson-Walter K J, Major M M, Johnson C K, et al. Examining the efficacy of decoy-based and psychological cyber deception[C] // Proc of USENIX Security Symp. Berkeley, CA: USENIX, 2021: 1127−1144

[22] Ferguson-Walter K J, Major M M, Johnson C K, et al. Cyber expert feedback: Experiences, expectations, and opinions about cyber deception[J]. Computers & Security, 2023, 130: 103268

[23] Jia Zhaopeng, Fang Binxing, Liu Chaoge, et al. Survey on cyber deception[J]. Journal on Communications, 2017, 38(12): 128−143(in Chinese)
(贾召鹏, 方滨兴, 刘潮歌, 等. 网络欺骗技术综述[J]. 通信学报, 2017, 38(12): 128−143)

[24] Zhang Li, Thing V L. Three decades of deception techniques in active cyber defense-retrospect and outlook[J]. Computers & Security, 2021, 106: 102288

[25] Hu Yongjin, Ma Jun, Guo Yuanbo. Research on network deception based on game theory[J]. Journal of Communications, 2018, 39(S2): 9−18(in Chinese)
(胡永进, 马骏, 郭渊博. 基于博弈论的网络欺骗研究[J]. 通信学报, 2018, 39(S2): 9−18)

[26] Zhu Mu, Anwar A H, Wan Zelin, et al. A survey of defensive deception: Approaches using game theory and machine learning[J]. IEEE Communications Surveys & Tutorials, 2021, 23(4): 2460−2493

[27] Kasneci E, Seßler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education[J]. Learning and Individual Differences, 2023, 103: 102274

[28] Kocoń J, Cichecki I, Kaszyca O, et al. ChatGPT: Jack of all trades, master of none[J]. Information Fusion, 2023, 99: 101861

[29] Zhou Ming, Duan Nan, Liu Shujie, et al. Progress in neural NLP: Modeling, learning, and reasoning[J]. Engineering, 2020, 6(3): 275−290

[30] Steingartner W, Galinec D, Kozina A. Threat defense: Cyber deception approach and education for resilience in hybrid threats model[J]. Symmetry, 2021, 13(4): 597

[31] Ziaie Tabari A, Ou Xinming. A multi-phased multi-faceted IoT honeypot ecosystem[C] //Proc of the 2020 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2020: 2121−2123

[32] Zarca A M, Bernabe J B, Skarmeta A, et al. Virtual IoT HoneyNets to mitigate cyberattacks in SDN/NFV-enabled IoT networks[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(6): 1262−1277

[33] Zhang Weizhe, Zhang Bin, Zhou Ying, et al. An IoT honeynet based on multiport honeypots for capturing IoT attacks[J]. IEEE Internet of Things Journal, 2019, 7(5): 3991−3999

[34] Srinivasa S, Pedersen J M, Vasilomanolakis E. Towards systematic honeytoken fingerprinting[C] //Proc of 13th Int Conf on Security of Information and Networks. New York: ACM , 2020: 1−5

[35] Reti D, Angeli T, Schotten H D. Honey Infiltrator: Injecting Honeytoken using netfilter[C] //Proc of the 2023 IEEE European Symp on Security and Privacy Workshops. Piscataway, NJ: IEEE, 2023: 465−469

[36] Tan Jinglei, Jin Hui, Hu Hao, et al. WF-MTD: Evolutionary decision method for moving target defense based on wright-fisher process[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 20(6): 4719−4732

[37] Qian Yaguan, Guo Yankai, Shao Qiqi, et al. EI-MTD: Moving target defense for edge intelligence against adversarial attacks[J]. ACM Transactions on Privacy and Security, 2022, 25(3): 1−24

[38] Javadpour A, Ja'fari F, Taleb T, et al. SCEMA: An SDN-oriented

cost-effective edge-based MTD approach[J]. IEEE Transactions on Information Forensics and Security, 2022, 18: 667−682

[39] Simmons C B, Shiva S G, Bedi H, et al. ADAPT: A game inspired attack-defense and performance metric Taxonomy[C] //Proc of the 28th IFIP TC11 Int Conf. Berlin: Springer, 2013: 344−365

[40] Liu Shouzhou, Shao Chengwu, Li Yanfu, et al. Game attack–defense graph approach for modeling and analysis of cyberattacks and defenses in local metering system[J]. IEEE Transactions on Automation Science and Engineering, 2021, 19(3): 2607−19

[41] Zhou Yuyang, Cheng Guang, Yu Shui. An SDN-enabled proactive defense framework for DDoS mitigation in IoT networks[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 5366−5380

[42] Khan M S, Siddiqui S, Ferens K. A cognitive and concurrent cyber kill chain model[G]//Computer and Network Security Essentials. Berlin: Springer, 2018: 585−602

[43] Warner C. Online Operations Kill Chain in CTI[EB/OL]. (2023-11-07)[2024-01-10]. https://warnerchad.medium.com/online-operations-kill-chain-in-cti-8b3c99848250

[44] Xiong Wenjun, Legrand E, Åberg O, et al. Cyber security threat modeling based on the MITRE Enterprise ATT&CK Matrix[J]. Software and Systems Modeling, 2022, 21(1): 157−177

[45] Webb J, Hume D. Campus IoT collaboration and governance using the NIST cybersecurity framework[C] //Proc of Living in the Internet of Things: Cybersecurity of the IoT-2018. London: IET, 2018: 1−7

[46] Muckin M, Fitch S C. A threat-driven approach to cyber security[R]. Washington: Lockheed Martin Corporation, 2014

[47] Akbar K A, Rahman F I, Singhal A, et al. The design and application of a unified ontology for cyber security[C] //Proc of Int Conf on Information Systems Security. Berlin: Springer, 2023: 23−41

[48] Underbrink A J. Effective cyber deception[G]//Cyber Deception: Building the Scientific Foundation. Berlin: Springer, 2016: 115−147

[49] Kaynar K. A taxonomy for attack graph generation and usage in network security[J]. Journal of Information Security and Applications, 2016, 29: 27−56

[50] Ye Yun, Xu Xishan, Qi Zhichang, et al. Attack graph generation algorithm for large-scale network system[J]. Journal of Computer Research and Development, 2013, 50(10): 2133−2139 (in Chinese)
(叶云, 徐锡山, 齐治昌, 等. 大规模网络中攻击图自动构建算法研究[J]. 计算机研究与发展, 2013, 50(10): 2133−2139)

[51] Ye Ziwei, Guo Yuanbo, Wang Chendong, et al. Survey on application of attack graph technology[J]. Journal on Communications, 2017, 38(11): 121−132 (in Chinese)
(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)

[52] Muñoz-González L, Sgandurra D, Barrère M, et al. Exact inference techniques for the analysis of Bayesian attack graphs[J]. IEEE Transactions on Dependable and Secure Computing, 2017, 16(2): 231−244

[53] Nadeem A, Verwer S, Moskal S, et al. Alert-driven attack graph generation using S-PDFA[J]. IEEE Transactions on Dependable and Secure Computing, 2021, 19(2): 731−746

[54] Durkota K, Lisý V, Bošanský B, et al. Hardening networks against strategic attackers using attack graph games[J]. Computers & Security, 2019, 87: 101578

[55] Wang Binghui, Gong N Z. Attacking graph-based classification via manipulating the graph structure[C] //Proc of the 2019 ACM SIGSAC Conf on Computer and Communications Security. New York: ACM, 2019: 2023−2040

[56] Jorjani M, Seifi H, Varjani A Y. A graph theory-based approach to detect false data injection attacks in power system AC state estimation[J]. IEEE Transactions on Industrial Informatics, 2020, 17(4): 2465−2475

[57] Naik N, Grace P, Jenkins P, et al. An evaluation of potential attack surfaces based on attack tree modelling and risk matrix applied to self-sovereign identity[J]. Computers & Security, 2022, 120: 102808

[58] Shin G Y, Hong S S, Lee J S, et al. Network security node-edge scoring system using attack graph based on vulnerability correlation[J]. Applied Sciences, 2022, 12(14): 6852

[59] Almohri H M J, Watson L T, Yao D, et al. Security optimization of dynamic networks with probabilistic graph modeling and linear programming[J]. IEEE Transactions on Dependable and Secure Computing, 2015, 13(4): 474−487

[60] Wang L, Jajodia S, Singhal A, et al. Security Risk Analysis of Enterprise Networks Using Probabilistic Attack Graphs[M]. Berlin: Springer, 2017

[61] Ou Xinming, Govindavajhala S, Appel A W. MulVAL: A logic-based network security analyzer[C] //Proc of USENIX Security Symp. Berkeley, CA: USENIX, 2005, 8: 113−128

[62] Chen Xiaojun, Fang Binxing, Tan Qingfeng, et al. Research on internal attack intent inference algorithm based on probabilistic attack graph[J]. Chinese Journal of Computers, 2014, 37(1): 62−72 (in Chinese)
(陈小军, 方滨兴, 谭庆丰, 等. 基于概率攻击图的内部攻击意图推断算法研究[J]. 计算机学报, 2014, 37(1): 62−72)

[63] Sun Xiaoyan, Dai Jun, Liu Peng, et al. Using Bayesian networks for probabilistic identification of zero-day attack paths[J]. IEEE Transactions on Information Forensics and Security, 2018, 13(10): 2506−2521

[64] Sahu A, Davis K. Structural learning techniques for Bayesian attack graphs in cyber physical power systems[C] //Proc of 2021 IEEE Texas Power and Energy Conf (TPEC). Piscataway, NJ: IEEE, 2021: 1−6

[65] Matthews I, Soudjani S, van Moorsel A. Stochastic simulation techniques for inference and sensitivity analysis of Bayesian attack graphs[C] //Proc of Int Conf on Science of Cyber Security. Berlin: Springer, 2021: 171−186

[66] Asvija B, Eswari R, Bijoy M B. Bayesian attack graphs for platform virtualized infrastructures in clouds[J]. Journal of Information

Security and Applications, 2020, 51: 102455

[67] Anwar A H, Kamhoua C A. Cyber deception using honeypot allocation and diversity: A game theoretic approach[C] //Proc of the 19th Annual Consumer Communications & Networking Conf (CCNC). Piscataway, NJ: IEEE, 2022: 543−549

[68] Li Shuai, Wang Ting, Ma Ji, et al. A three-party attack-defense deception game model based on evolutionary[C] //Proc of the 3rd Int Conf on Consumer Electronics and Computer Engineering (ICCECE). Piscataway, NJ: IEEE, 2023: 51−56

[69] Zhou Yuyang, Cheng Guang, Jiang Shanqing, et al. Cost-effective moving target defense against DDoS attacks using trilateral game and multi-objective Markov decision processes[J]. Computers & Security, 2020, 97: 101976

[70] Thakoor O, Tambe M, Vayanos P, et al. General-sum cyber deception games under partial attacker valuation information[C] //Proc of Int Foundation for Autonomous Agents and Multiagent Systems (AAMAS). Richland, SC: AAMAS, 2019: 2215−2217

[71] Liu Jieling, Wang Zhiliang, Yang Jiahai, et al. Deception maze: A stackelberg game-theoretic defense mechanism for intranet threats[C] // Proc of IEEE Int Conf on Communications (ICC 2021). Piscataway, NJ: IEEE, 2021: 1−6

[72] Sayed M A, Anwar A H, Kiekintveld C, et al. Cyber deception against zero-day attacks: A game theoretic approach[C] //Proc of Int Conf on Decision and Game Theory for Security. Berlin: Springer, 2022: 44−63

[73] Wahab O A, Bentahar J, Otrok H, et al. Resource-aware detection and defense system against multi-type attacks in the cloud: Repeated Bayesian stackelberg game[J]. IEEE Transactions on Dependable and Secure Computing, 2019, 18(2): 605−622

[74] Sengupta S, Chowdhary A, Huang Dijiang, et al. General sum Markov games for strategic detection of advanced persistent threats using moving target defense in cloud networks[C] //Proc of Decision and Game Theory for Security: 10th Int Conf. Berlin: Springer, 2019: 492−512

[75] Sengupta S, Chowdhary A, Huang Dijiang, et al. Moving target defense for the placement of intrusion detection systems in the cloud[C] //Proc of Decision and Game Theory for Security: 9th Int Conf. Berlin: Springer, 2018: 326−345

[76] Huang Linan, Zhu Quanyan. Dynamic Bayesian games for adversarial and defensive cyber deception[G]//Autonomous Cyber Deception: Reasoning, Adaptive Planning, and Evaluation of HoneyThings. Berlin: Springer, 2019: 75−97

[77] Yang Junnan, Zhang Hongqi, Zhang Chuanfu. Network defense decision-making method based on stochastic game and improved WoLF-PHC[J]. Journal of Computer Research and Development, 2019, 56(5): 942−954 (in Chinese)
(杨峻楠, 张红旗, 张传富. 基于随机博弈与改进 WoLF-PHC 的网络防御决策方法[J]. 计算机研究与发展, 2019, 56(5): 942−954)

[78] Tsemogne O, Hayel Y, Kamhoua C, et al. Game-theoretic modeling of cyber deception against epidemic botnets in Internet of things[J]. IEEE Internet of Things Journal, 2021, 9(4): 2678−2687

[79] Thakoor O, Tambe M, Vayanos P, et al. Cyber camouflage games for strategic deception[C] //Proc of Decision and Game Theory for Security: 10th Int Conf. Berlin: Springer, 2019: 525−541

[80] Shinde A, Doshi P, Setayeshfar O. Cyber attack intent recognition and active deception using factored interactive POMDPs[C] //Proc of the 20th Int Conf on Autonomous Agents and MultiAgent Systems. Richland, SC: AAMAS, 2021: 1200−1208

[81] Zhang Tao, Xu Changqiao, Shen Jiahao, et al. How to disturb network reconnaissance: A moving target defense approach based on deep reinforcement learning[J]. IEEE Transactions on Information Forensics and Security, 2023, 18: 5735-5748

[82] Tian Wen, Du Miao, Ji Xiaopeng, et al. Honeypot detection strategy against advanced persistent threats in industrial internet of things: A prospect theoretic game[J]. IEEE Internet of Things Journal, 2021, 8(24): 17372−17381

[83] Yoon S, Cho J H, Kim D S, et al. Attack graph-based moving target defense in software-defined networks[J]. IEEE Transactions on Network and Service Management, 2020, 17(3): 1653−1668

[84] Hu Chenao, Yan Xuefeng. Dynamic trilateral game model for attack graph security game[C] //Proc of IOP Conf Series: Materials Science and Engineering. Bristol: IOP Publishing, 2020, 790: 012112

[85] Anwar A H, Kamhoua C. Game theory on attack graph for cyber deception[C] //Proc of Int Conf on Decision and Game Theory for Security. Berlin: Springer, 2020: 445−456

[86] Milani S, Shen W, Chan K S, et al. Harnessing the power of deception in attack graph-based security games[C] //Proc of Decision and Game Theory for Security: 11th Int Conf. Berlin: Springer, 2020: 147−167

[87] Wu Hua, Gu Yu, Cheng Guang, et al. Effectiveness evaluation method for cyber deception based on dynamic bayesian attack graph[C] //Proc of the 3rd Int Conf on Computer Science and Software Engineering. New York: ACM, 2020: 1−9

[88] Huang Weigui, Sun Yifeng, Ou Wang, et al. A flow scheduling model for SDN Honeypot using multi-layer attack graphs and signaling game[C] //Proc of 2021 7th Int Conf on Computer and Communications (ICCC). Piscataway, NJ: IEEE, 2021: 2012−2020

[89] Buczkowski P, Malacaria P, Hankin C, et al. Optimal security hardening over a probabilistic attack graph: A case study of an industrial control system using CySecTool[C] //Proc of the 2022 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems. New York: ACM, 2022: 21−30

[90] Outkin A V, Schulz P V, Schulz T, et al. Defender policy evaluation and resource allocation with MITRE ATT&CK evaluations data[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 20(3): 1909−1926

[91] Guo Mingyu, Li Jialiang, Neumann A, et al. Practical fixed-parameter algorithms for defending active directory style attack graphs[C] //Proc of the AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2022, 36(9): 9360−9367

[92] Ma Haoxiang, Han Shuo, Leslie N, et al. Optimal decoy resource allocation for proactive defense in probabilistic attack graphs[C]//Proc of the 2023 Int Conf on Autonomous Agents and Multiagent Systems. Richland, SC: AAMAS, 2023: 2616−2618

[93] Li Lening, Ma Haoxiang, Han Shuo, et al. Synthesis of proactive sensor placement in probabilistic attack graphs[C]//Proc of the 2023 American Control Conf (ACC). Piscataway, NJ: IEEE, 2023: 3415−3421

[94] Cao Yihan, Li Siyu, Liu Yixin, et al. A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to chatgpt[J]. arXiv preprint, arXiv: 2303.04226, 2023

[95] Ziems N, Yu Wenhao, Zhang Zhihan, et al. Large language models are built-in autoregressive search engines[J]. arXiv preprint, arXiv: 2305. 09612, 2023

[96] Panda S. Enhancing PDF interaction for a more engaging user experience in library: Introducing ChatPDF[J]. IP Indian Journal of Library Science and Information Technology, 2023, 8(1): 20−25

[97] Firat M, Kuleli S. What if GPT4 became autonomous: The Auto-GPT project and use cases[J]. Journal of Emerging Computer Technologies, 2023, 3(1): 1−6

[98] Deng Gelei, Liu Yi, Mayoral-Vilches V, et al. PentestGPT: An LLM-empowered automatic penetration testing tool[J]. arXiv preprint, arXiv: 2308.06782, 2023

[99] Renaud K, Warkentin M, Westerman G. From ChatGPT to HackGPT: Meeting the Cybersecurity Threat of Generative AI[M]. Cambridge, MA: MIT Sloan Management Review, 2023

[100] Aleena N. Large Language Models in Cybersecurity: Upcoming AI Trends in 2023-24 [EB/OL]. 2023[2024-01-10]. https://hubs.ly/Q01XQM5q0

[101] Das S S, Dutta A, Purohit S, et al. Towards automatic mapping of vulnerabilities to attack patterns using large language models[C]//Proc of the 2022 IEEE Int Symp on Technologies for Homeland Security (HST). Piscataway, NJ: IEEE, 2022: 1−7

[102] Tian Zhihong, Fang Binxing, Liao Qing, et al. Cybersecurity assurance system in the new era and development suggestions thereof: From self-defense to guard[J]. Strategic Study of CAE, 2023, 25(6): 96−105 (in Chinese)
(田志宏, 方滨兴, 廖清, 等. 从自卫到护卫: 新时期网络安全保障体系构建与发展建议[J]. 中国工程科学, 2023, 25(6): 96−105)

Wang Rui, born in 1994. PhD candidate. His main research interests include network security, attack-defense game and deception defense.

王　瑞, 1994 年生. 博士研究生. 主要研究方向为网络安全、攻防博弈、欺骗防御.

Yang Changjiang, born in 2000. Master candidate. His main research interests include network security and deception defense.

阳长江, 2000 年生. 硕士研究生. 主要研究方向为网络安全、欺骗防御.

Deng Xiangdong, born in 2000. Master candidate. His main research interests include network security and deception defense.

邓向东, 1998 年生. 硕士研究生. 主要研究方向为网络安全、欺骗防御.

Liu Yuan, born in 1986. PhD, professor, PhD supervisor. Distinguished member of CCF. Her main research interests include network security, mechanism design, and game theory.

刘　园, 1986 年生. 博士, 教授, 博士生导师. CCF 杰出会员. 主要研究方向为网络安全、机制设计、博弈理论.

Tian Zhihong, born in 1978. PhD, professor, PhD supervisor. Distinguished member of CCF. His research interests include network attack and defense confrontation, APT detection and traceability, and industrial control security.

田志宏, 1978 年生. 博士, 教授, 博士生导师. CCF 杰出会员. 主要研究方向为网络攻防对抗、APT 检测与溯源、工控安全.