软件学报 ISSN 1000-9825, CODEN RUXUEW
[doi: 10.13328/j.cnki.jos.007162] [CSTR: 32375.14.jos.007162]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

面向 APT 攻击的溯源和推理研究综述*

杨秀璋[1,2,3], 彭国军[1,2], 刘思德[1,2], 田 杨[1,2], 李晨光[1,2], 傅建明[1,2]

[1](武汉大学 国家网络安全学院, 湖北 武汉 430072)

[2](空天信息安全与可信计算教育部重点实验室 (武汉大学 国家网络安全学院), 湖北 武汉 430072)

[3](贵州大学 贵州省大数据产业发展应用研究院, 贵州 贵阳 550025)

通信作者: 彭国军, E-mail: guojpeng@whu.edu.cn

摘 要: 高级可持续性威胁 (advanced persistent threat, APT) 是一种新型网络攻击, 具有极强的组织性、隐蔽性、持续性、对抗性和破坏性, 给全球网络安全带来严重危害. 传统 APT 攻击防御倾向于构建模型检测攻击的恶意性或识别家族类别, 以被动防御为主, 缺乏全面及深入地梳理 APT 攻击溯源和推理领域的工作. 基于此, 围绕 APT 攻击的溯源和推理的智能化方法开展综述性研究. 首先, 提出 APT 攻击防御链, 有效地将 APT 攻击检测、溯源和推理进行区分和关联; 其次, 详细比较 APT 攻击检测 4 个任务的相关工作; 然后, 系统总结面向区域、组织、攻击者、地址和攻击模型的 APT 攻击溯源工作; 再次, 将 APT 攻击推理划分为攻击意图推理、攻击路径感知、攻击场景还原、攻击阻断和反制这 4 个方面, 对相关研究进行详细总结和对比; 最后, 讨论 APT 攻击防御领域的热点主题、发展趋势和挑战.

关键词: 高级可持续威胁; 网络安全; 攻击溯源; 攻击推理; 人工智能

中图法分类号: TP393

中文引用格式: 杨秀璋, 彭国军, 刘思德, 田杨, 李晨光, 傅建明. 面向APT攻击的溯源和推理研究综述. 软件学报. http://www.jos.org.cn/1000-9825/7162.htm

英文引用格式: Yang XZ, Peng GJ, Liu SD, Tian Y, Li CG, Fu JM. Survey on Attribution and Inference Research for APT Attacks. Ruan Jian Xue Bao/Journal of Software (in Chinese). http://www.jos.org.cn/1000-9825/7162.htm

Survey on Attribution and Inference Research for APT Attacks

YANG Xiu-Zhang[1,2,3], PENG Guo-Jun[1,2], LIU Si-De[1,2], TIAN Yang[1,2], LI Chen-Guang[1,2], FU Jian-Ming[1,2]

[1](School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China)

[2](Key Laboratory of Aerospace Information Security and Trusted Computing of Ministry of Education (School of Cyber Science and Engineering, Wuhan University), Wuhan 430072, China)

[3](Guizhou Big Data Academy, Guizhou University, Guiyang 550025, China)

Abstract: Advanced persistent threat (APT) is a novel form of cyberattack that is well-organized, stealthy, persistent, adversarial, and destructive, resulting in catastrophic consequences for global network security. Traditional APT attack defenses tend to construct models to detect whether the attacks are malicious or identify the malicious family categories, primarily employing a passive defense strategy and lacking comprehensive and in-depth exploration of the field of APT attack attribution and inference. In light of this, this study focuses on the intelligent methods of APT attack attribution and inference to conduct a survey study. Firstly, an overall defense chain framework for APT attacks is proposed, which can effectively distinguish and correlate APT attack detection, attribution, and inference. Secondly, the work related to the four tasks of APT attack detection is reviewed in detail. Thirdly, APT attack attribution research is systematically summarized for regions, organizations, attackers, addresses, and attack models. Then, APT attack inference is divided into four aspects: attack intent inference, attack path perception, attack scenario reconstruction, and attack blocking and countermeasures, and relevant works are summarized and compared in detail. Finally, the hot topics, development trends, and challenges in the field of APT attack defense are discussed.

* 基金项目: 国家自然科学基金 (62172308, 61972297, 62172144, U1636107, 62062019)
收稿时间: 2023-02-12; 修改时间: 2023-05-10, 2023-11-18; 采用时间: 2024-01-30; jos 在线出版时间: 2024-10-23

Key words: advanced persistent threat (APT); network security; attack attribution; attack inference; artificial intelligence

随着全球网络形势日益复杂, 国家之间的网络对抗正变得越来越激烈, 我国的网络安全形势十分严峻. 高级可持续性威胁 (advanced persistent threat, APT) 攻击是利用高级入侵手段并针对特定目标实施可持续的新型网络攻击, 旨在刺探、收集和窃取核心情报, 并监控、渗透和破坏关键基础设施及网络设备[1−3]. APT 攻击通常由特定民族或国家支持, 由于 APT 攻击具有极强的组织性、隐蔽性、持续性、对抗性和破坏性, 它给全球的网络安全带来严重的危害[4,5]. 2010 年, 伊朗布什尔核电站设备遭受震网 (Stuxnet) 蠕虫攻击, 高度复杂的恶意代码及多个 0day 漏洞造成约 1 000 台铀浓缩离心机故障[6]. 2015 年 12 月 23 日, 乌克兰伊万诺弗兰科夫斯克地区遭受 APT 攻击导致大规模停电, 该攻击是由于电力系统感染携带 Killdisk 组件的恶意木马 BlackEnergy 所致[7]. 2020 年 12 月 13 日, FireEye 公司发布报告, 声称 SolarWinds 旗下的软件遭到供应链攻击, 黑客通过篡改源代码并添加后门实施网络攻击[8]. 2022 年 9 月, 根据 360 公司最新溯源研究披露, 美国国家安全局 (NSA) 针对某大学发起上千次网络攻击活动, 通过精心构造的武器库控制内网设备并窃取高价值数据, 再利用 49 台跳板机和多个 0day 漏洞 (譬如 Extremeparr 和 Ebbisland) 开展定向攻击, 造成巨大的损失[9]. 这些攻击事件表明, APT 攻击给全球政治、经济、科技、教育、文化各方面带来严重影响, 我国作为主要受害国正面临严峻的网络攻击威胁[3,10,11]. 由此可见, 如何快速精准地检测及溯源 APT 攻击已成为重要的研究主题, 并且针对 APT 攻击事件的检测、溯源和推理研究将成为国家之间网络博弈的重要技术手段.

传统 APT 攻击防护需要借助大量的专家知识和人工标注, 通过定制检测规则和特征库实现, 典型技术包括特征值检测技术、校验和检测技术、启发式检测技术和主动防御技术[12−14]. 然而, 该类方法自动化和智能化程度低, 过度依赖专家知识, 较难准确感知攻击行为和挖掘攻击意图, 面对大规模安全日志、网络流量和恶意样本时, 其准确率和鲁棒性较差, 且无法有效对抗具有混淆、伪装、逃逸和欺骗的新型高隐蔽 APT 攻击, 典型的高隐蔽攻击包括混淆恶意请求攻击[15]、离地攻击 (living-off-the-land attack)[16]、无文件攻击 (fileless attack)[17]、躲避检测的逃逸攻击 (evasion attack)[18]、基于匿名网络的隐蔽攻击[19]和利用 0day 漏洞的网络攻击[20]. 近些年来, 以机器学习、神经网络和知识图谱[21]为代表的人工智能技术正在蓬勃发展, 并且在各领域取得重要进步, 智能化的方法和系统有效提高生产力, 增强自主决策和推理能力. 人工智能技术被广泛应用于安全领域, 典型应用包括网络入侵检测[22]、恶意软件检测[12]、攻击行为预测[23]、工业控制系统安全防护[24]以及个人隐私保护[25]. 同样, 利用人工智能技术实现安全防护和 APT 攻击检测、溯源和推理已成为重要研究方向并取得一定的进展, 这类方法能更好地检测攻击行为、溯源攻击来源和推理攻击意图, 将帮助安全分析人员及时掌握网络安全态势.

(1) 在攻击检测方面, 其基本思路是通过提取关键特征并构建智能模型来识别恶意网络攻击及其行为. 其中, Liras 等人[26]构建 4 种机器学习模型来学习所提取的静态和动态特征并有效识别 APT 攻击事件中的恶意软件. Downing 等人[27]提出 DeepReflect 系统, 旨在构建无监督的神经网络定位和识别二进制文件中的恶意软件组件及函数行为. Li 等人[28]构建融合注意力的图神经网络 (graph neural network, GNN) 模型来学习审计日志的节点属性并增强边缘, 从而实现对 APT 攻击的识别. Han 等人[29]设计一种基于溯源图 (provenance graph) 的运行时 APT 检测系统, 旨在利用丰富的上下文信息和较少的先验知识来识别 APT 攻击.

(2) 在攻击溯源方面, 其基本思路是结合 APT 攻击的特点提取溯源特征, 再构建智能模型以识别和挖掘攻击活动中的线索和痕迹, 最终实现对 APT 攻击的溯源追踪. FireEye 公司[30]结合 APT 组织特点生成文档主题摘要, 利用相似性度量和聚类算法实现 APT 组织溯源. Alrabaee 等人[31]详细分析二进制文件在编译过程中遗留的特征, 再利用支持向量机 (support vector machine, SVM) 算法来溯源恶意二进制文件的作者. 李昂[32]提出一种基于恶意代码基因的攻击组织特征提取方法, 构建融合自注意力机制的循环神经网络 (recurrent neural network, RNN) 模型识别 APT 组织. 黄克振等人[33]利用攻击事件线索和情报特征来构建溯源关系图谱并追踪 APT 组织.

(3) 在攻击推理方面, 其基本思路是利用人工智能技术来实现攻击意图推理、攻击路径感知、攻击场景还原以及攻击阻断和反制. Hossain 等人[34]首次利用溯源图和因果依赖关系重构 APT 攻击并还原攻击场景. Xiong 等人[35]提出一种基于状态跟踪和检测的 Conan 框架, 利用有限状态自动机来识别审计日志的状态信息和语义知识,

再构建智能策略来重构 APT 攻击链. Zimba 等人[36]通过贝叶斯网络 (Bayesian network) 来识别云计算中 APT 攻击的最短路径并模拟攻击场景. 陈瑞东等人[5]通过研究 APT 经典案例, 详细分析 APT 攻击检测与反制技术并提出 APT 整体防御方案.

此外, 针对 APT 攻击的综述性文章陆续出现. Lemay 等人[37]系统归纳和分析了 APT 攻击组织及其活动的相关知识, 涵盖全球多个地区约 40 个 APT 组织的技战术特点. 文献 [12] 和文献 [23] 详细总结恶意软件智能检测领域, 囊括特征提取、特征处理和分类器 3 个关键步骤. Alshamrani 等人[10]详细综述 APT 攻击不同阶段的方法、技术和应用, 以及 APT 攻击智能检测和缓解方案. Stojanović 等人[4]聚焦于 APT 数据集和攻击模型, 对 APT 检测相关的文献进行综述. Talib 等人[38]针对 APT 攻击的命令和控制 (command and control, C&C) 阶段开展综述研究. 刘潮歌等人[39]详细归纳定向网络攻击追踪溯源层次化模型, 并以该模型为基础建立包含“欺骗环境构建”“多源线索提取”“线索分析挖掘”的追踪溯源纵深体系. 宋文纳等人[14]系统综述学术界和产业界在恶意代码溯源领域的研究工作. 付钰等人[3]基于大数据分析总结 APT 攻击检测在网络攻击流量异常检测、恶意代码异常检测、社交网络安全事件挖掘和安全事件关联分析 4 个领域的研究进展. 潘亚峰等人[40]面向 APT 攻击的场景重构方法进行综述, 归纳了基于经验知识、因果关系、语义相似性和机器学习的方法和案例. Chen 等人[41]总结和归纳基于机器学习的物联网中的高级可持续性威胁检测方法.

然而, 尽管已经存在上述文献综述和相关研究, 但大多数综述倾向于单一领域的具体应用, 更多的研究聚焦于 APT 攻击检测, 且仍然缺乏对 APT 攻击开展深入、系统、全面地梳理和总结. 此外, APT 攻击通常涉及多个阶段, APT 安全防护是一个系统的过程, 现有工作尚无将 APT 攻击检测、溯源和推理相关联的深入研究和总结, 譬如文献 [10,12,23] 聚焦于 APT 攻击检测研究, 文献 [14,33,39] 关注于 APT 攻击的溯源和追踪, 文献 [4,37,38,41] 仅从单一视图 (即数据集、APT 组织、命令与控制、物联网中的威胁事件) 对 APT 攻击开展研究, 并且现有文献较少对 APT 攻击溯源和推理进行梳理. 同时, 传统研究和综述更倾向于规则匹配和机器学习方法, 而面对 APT 攻击产生的大规模审计日志、网络流量和终端样本, 较难快速精准地识别恶意攻击、溯源攻击来源及推理攻击意图, 如何将深度学习、溯源图和知识图谱等智能化方法有效结合将变得至关重要.

为此, 本文对面向 APT 攻击的溯源和推理方法开展系统、深入、全面地归纳和总结, 结合 APT 生命周期详细分析和对比各类工作, 并总结 APT 攻击防御的研究趋势和挑战, 展望未来的研究方向. 本文的主要贡献如下.

(1) 本文以 APT 攻击生命周期为主线, 提出 APT 攻击防御链, 分别对 APT 攻击溯源和推理的智能化方法进行全面地总结、关联、归纳和分析. 在 APT 攻击溯源方面, 给出具体的定义, 并详细总结面向区域、组织、攻击者、地址和攻击模式的溯源工作; 在 APT 攻击推理方面, 本文将其划分为攻击意图推理、攻击路径感知、攻击场景还原、攻击阻断和反制 4 个层面, 并深入分析和比较现有智能化推理工作.

(2) 本文系统地回顾 APT 攻击溯源和推理相关研究工作, 全面概述不同智能化方法的实现过程以及设计特点, 探索和对比学术界和产业界相关工作的优缺点, 并展望 APT 攻击防御的研究趋势及存在挑战. 此外, 本文针对溯源对抗和推理对抗进行分析和思考.

(3) 本文实现一个更全面、更有层次、更细粒度、更多维的综述研究, 旨在从前沿和智能视角分析网络安全现状, 给出将检测、溯源和推理相结合的 APT 攻击安全防护建议, 并总结相关发现和各阶段存在的问题及困难.

(4) 通过系统地总结和分析, 本文将尝试回答 APT 攻击领域的 4 类研究问题: ① APT 攻击的技战术特点和整体防御技术是什么? 如何有效地将 APT 攻击检测、溯源及推理系统关联? ② APT 攻击溯源和推理现有方法的特点是什么? 这些方法和策略的基本思路如何? 各自具有哪些优点和缺陷? ③ 如何将人工智能技术应用于 APT 攻击防护领域, 是否能识别高隐蔽、高对抗、高威胁的 APT 攻击? ④ APT 攻击未来的研究趋势及面临的挑战是什么? 如何解决现有智能方法的局限性?

需要注意, 本文研究以被动防御模式为主, 旨在将智能化技术应用于 APT 攻击检测、溯源及推理任务, 结合 APT 防御链开展被动防御工作. 然而, 被动防御存在天生的局限性, 其无法及时构建完整的证据链, 难以应对复杂多变的网络攻击, 无法满足当代网络环境下对 APT 攻击的实时精准溯源以及将攻击阻断在源头的需求.

1 APT 攻击特点及防御概述

当前, 国内外学者和安全人员围绕 APT 攻击已经开展不少研究并取得一定成果. 本文首先介绍 APT 攻击的特点, 并围绕 APT 攻击生命周期详细概述常见的防御方法, 最终引出 APT 攻击防御链.

1.1 APT 攻击特点与生命周期

APT 攻击旨在针对特定目标实施隐蔽且持久的网络入侵, 实现对目标的控制和破坏, 以及捕获关键数据和资产. “Advanced Persistent Threat (APT)”术语最早由美国空军在 2006 年创造, 用以关注具有针对性的入侵活动[1−3,11]. 根据 360 公司最新发布的 APT 研究报告显示[42], 我国是 APT 攻击的主要受害者, 针对我国的攻击来源以南亚、东南亚、东亚以及欧美地区的 APT 组织最为活跃, 被攻击的行业包括国防军工、金融、医疗、科研、教育和高新技术等. 表 1 展示部分长期活跃并针对中国大陆实施攻击活动的 APT 组织及攻击行为信息[43].

表 1 APT 组织及攻击行为信息

序号	组织编号	组织名称	最早发现时间	攻击行业	常用攻击手法
1	APT32	海莲花	2012年	政府、ICT供应商、教育	鱼叉攻击、水坑攻击
2	APT-C-01	毒云藤	2007年	政府、科研、国防	鱼叉攻击、水坑攻击
3	APT-C-08	蔓灵花	2013年	工业、军工、政府	漏洞攻击、鱼叉攻击
4	APT-C-06	Darkhotel	2010年	贸易、军工、科研	漏洞攻击、鱼叉攻击
5	APT-C-09	摩诃草	2009年	政府、军工、基础设施	社工攻击、鱼叉攻击
6	APT-C-55	Kimsuky	2013年	政府、国防、教育	鱼叉攻击、水坑攻击
7	APT-C-26	Lazarus	2007年	政府、企业、数字货币	鱼叉攻击、恶意代码
8	APT-C-24	响尾蛇	2012年	政府、军事、医疗	鱼叉攻击、水坑攻击
9	APT-C-40	方程式	2015年	政府、军工、工业制造	漏洞攻击、恶意代码
10	APT-C-12	蓝宝菇	2011年	政府、军工、航空航天	鱼叉攻击、水坑攻击

APT 攻击最典型的特点是高级性、持续性和高威胁性, 同时包括隐蔽性、模块化、对抗性和破坏性. 其中, 高级性表示 APT 攻击具有较高的组织性, 通常受国家背景支持, 攻击者由技术精湛的黑客组成, 会针对特定目标发起具有政治、军事或商业动机的网络攻击, 并且攻击方式隐蔽、作战意图明确, 甚至会利用多种漏洞和混淆对抗技术[44]来逃避各类安全机制的检测, 如 Stuxnet 事件[6]. 持续性表示 APT 攻击事件通常会在目标网络中潜伏较长时间, 攻击者会采用反复渗透、横向移动和优化攻击方式, 以实现对目标的持续控制并获得长期收益, 整个攻击时间跨度大且具有“low-and-slow”特点[4,10]. 据 FireEye 最新调查发现[45], 2020 年亚太地区 APT 攻击的中值停留时间超 76 天, 如何从海量信息中准确检测 APT 攻击尤为关键. 高威胁性表示 APT 攻击会针对特定目标和对象来设计, 通常会对国家机构、大型公司、关键基础设施、金融市场和重要部门等开展攻击, 攻击所造成的数据泄露、经济损失、系统破坏、行业崩溃等现象具有严重的危害性, 影响范围广泛. 由此可见, 针对我国面临日益严重的 APT 攻击活动, 如何提升 APT 攻击防御能力已成为我国网络空间安全保障的重要基础和急迫需求. 此外, APT 攻击通常具有相对固定的生命周期, 图 1 展示 APT 攻击的网络杀伤链模型[5,46].

在该模型中, APT 攻击的生命周期由 7 个关键阶段组成, 包括目标侦察、武器构建、载荷投递、漏洞利用、安装植入、命令与控制、任务执行. 目标侦察旨在获取目标系统信息, 搜集目标弱点知识, 并根据收集情报设计入侵路径, 典型技术如端口扫描、信息收集等; 武器构建旨在生成一个包含漏洞、后门或恶意代码, 并且可传输的武器载体, 典型技术如恶意代码、供应链后门等; 载荷投递旨在利用恶意链接、文件、邮件或设备等向目标投递定制的攻击载荷, 典型技术如钓鱼邮件、水坑攻击等; 漏洞利用旨在利用系统漏洞攻击受害者并执行相关的漏洞行为, 典型技术如 Web 漏洞、进程注入等; 安装植入是在目标指定位置安装恶意软件或木马, 典型技术如网络植入、横向移动等; 命令与控制旨在建立远程控制目标系统路径并对被控主机下达攻击指令, 典型技术如 C&C 通信、收集情报等; 任务执行是实施攻击任务、隐藏攻击痕迹并将远程数据回传以完成预期计划的过程, 典型技术如 C&C 回传、清除日志. 此后, ATT&CK (adversarial tactics, techniques, and common knowledge) 模型被提出, 它是一种攻击行为的知识

框架, 通过若干矩阵、战术、技术和结构化威胁信息来描述和评估 APT 攻击和威胁狩猎的攻防能力[43,47].

然而, 由于攻击目标、技术战术和攻击行为存在区别, APT 攻击在生命周期的不同阶段的表现和被检测、溯源及推理的难度也各不相同. 在攻击前期的目标侦察、武器构建、载荷投递中, 防御措施主要是对各种网络探测行为及武器和载荷对象的检测; 在攻击实施阶段的漏洞利用、安装植入、命令与控制中, 防御措施以主机层面对权限提升等行为的检测, 以及网络层面对命令控制通信、敏感数据泄露等行为的识别为主; 在攻击后续阶段的命令与控制、任务执行中, 防御措施将以审计日志和网络流量对比分析为主. 前中期检测难度相对较低, 但容易被混淆、对抗、加密、漏洞等技术影响, 从而降低检测的准确率; 后期检测和防御付出的代价较大, 实时检测和鲁棒性较差, 且日志痕迹存在被清除和规避的风险, 给 APT 攻击检测和取证造成严重影响. 为进一步突出 APT 攻击的隐蔽性和高威胁性, 后文图 2 展示了 Darkhotel 利用双星漏洞发起 APT 攻击的过程[48]. 该攻击通过投递包含漏洞文档的邮件来访问远程的 JS 文件, 再触发浏览器漏洞 (CVE-2020-0674 和 CVE-2019-17026) 启动 ShellCode 并下载远程木马实施持续性攻击. 由图 2 可知, 一次完整的 APT 攻击会涉及多个阶段且利用高隐蔽攻击技术, 其攻击痕迹会分布在多个日志中, 而现有的攻击检测与溯源方法存在较大不足, 无法有效还原攻击场景. 基于此, 本文将开展面向 APT 攻击的智能溯源和推理研究, 系统梳理和分析现有工作的优缺点, 探索和评估不同方法实现 APT 攻击防御的过程、特点及能力.

图 1 APT 攻击整体流程

1.2 APT 攻击防御链

针对现有 APT 攻击研究和文献综述仅从单一视角展开, 忽视 APT 攻击生命周期对安全防护的影响, 未有效地将 APT 攻击检测、溯源和推理关联, 缺乏深入系统的综述研究, 本文创新性提出 APT 攻击防御链框架. 首次将 APT 攻击检测、溯源和推理结合, 并系统地梳理应用人工智能技术实现 3 个关键阶段的现有研究工作, 力争从根源上遏制 APT 攻击. 整个防御链如图 3 所示, 结合 APT 攻击的生命周期和研究问题划分. 其中, 检测旨在发现攻击事件和识别攻击行为, 溯源旨在追踪和定位攻击来源, 推理旨在感知后续攻击行为、路径及意图.

(1) APT 攻击检测旨在判断网络攻击的恶意性、恶意家族, 以及识别恶意行为并定位关键代码. 基于此, 本文将 APT 攻击检测划分为恶意性检测、恶意家族检测、恶意行为识别和恶意代码定位, 其研究的问题分别是判断恶意攻击或良性访问、检测样本的恶意家族、识别攻击的恶意行为以及定位关键代码和功能函数区域. 该阶段的网络攻击已发生, 其核心思想是归纳已知的知识或挖掘潜在的关联特征来构建检测模型, 并实现对应的任务. APT 攻击检测的研究趋势包括检测粒度越来越细 (如定位恶意行为的关键代码)、先验知识越来越少 (如构建半监督学习模型), 甚至是实时检测未知网络攻击.

(2) APT 攻击溯源旨在挖掘网络攻击中的线索和痕迹以确定攻击者身份和地址的过程. 本文按溯源对象将 APT 攻击溯源划分为 5 类, 即面向区域、组织、地址、作者和攻击模式的溯源工作. 该阶段的网络攻击正在发生, 因此需利用已有的信息去发现和关联更多的攻击活动信息, 力争更准确地追踪攻击来源、识别攻击者身份, 而非

检测阶段的目标. APT 攻击溯源的研究趋势表现在智能化和自动化程度更高、借助大数据分析或知识图谱相关技术的研究更多, 并且更少依赖专家知识来构建溯源模型.

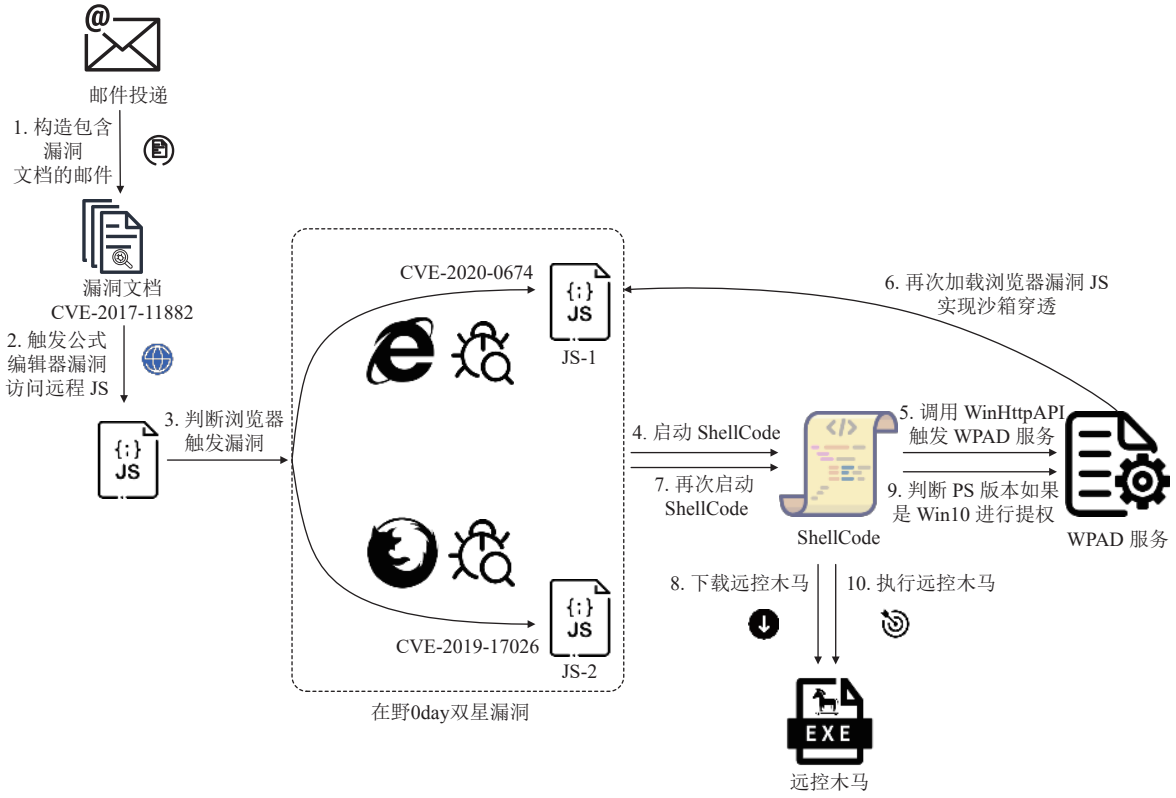

图 2 Darkhotel 利用双星漏洞发起 APT 攻击的过程

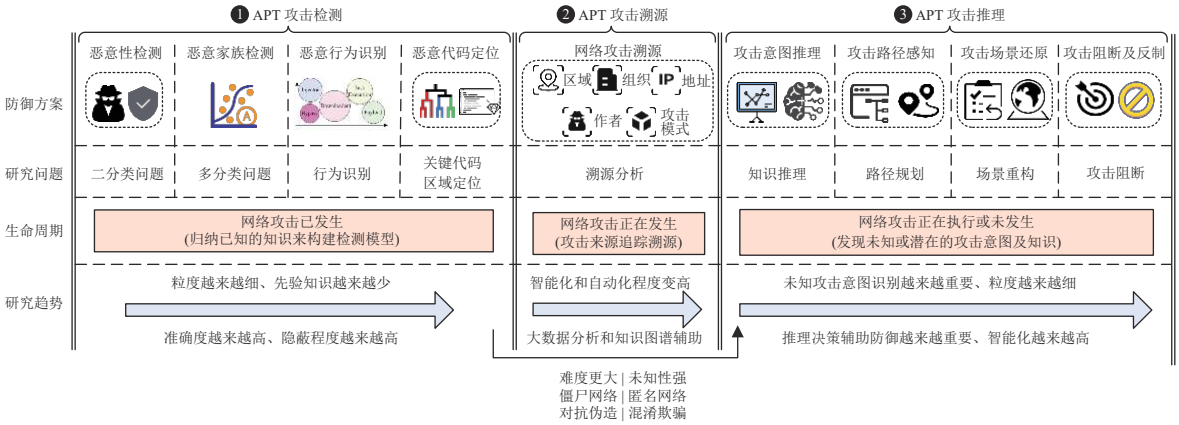

图 3 APT 攻击防御链

(3) APT 攻击推理旨在探索和预测攻击后续的意图或行为, 并尽可能地阻断攻击和反制对手. 本文通过对学术界和产业界相关工作的详细分析, 将 APT 攻击推理划分为 4 个关键任务, 包括攻击意图推理、攻击路径感知、攻击场景还原和攻击阻断及反制, 其研究的问题分别是攻击意图行为和决策的知识推理、攻击实施过程中路径规划及探测、整个攻击场景的重构, 以及结合攻击的意图、路径和场景实现 APT 攻击的阻断. 该阶段的网络攻击正在执行或未发生, 因此需要通过构建智能化算法或知识图谱来发现未知或潜在的攻击意图, 最终实现推理和反制. 尽

管 APT 攻击推理的研究相对较少, 但本文预测其将成为未来研究的重点, 整个 APT 研究的趋势倾向于识别攻击意图、还原攻击场景, 其推理特征的粒度更细、关联的知识更丰富, 且利用自主决策引擎或强化学习算法的需求会逐渐增加, 最终致力于自动生成包含攻击意图和阻断决策的方案, 从而实现对 APT 攻击的遏制和反制.

综上, 本文提出了 APT 攻击防御链, 不同于先前的研究, 本文不再局限于单一视角, 而是将 APT 攻击检测、溯源和推理紧密关联, 力争实现未知性更强和对抗性更高的 APT 攻击溯源和推理研究, 并最终形成一个系统全面的 APT 攻击智能防御综述, 该研究具有重要的学术价值和应用前景.

2 APT 攻击检测

由于 APT 攻击会对恶意代码进行加壳、混淆、加密或伪装等处理, 恶意软件数量逐年增加, 从而导致现有检测模型或系统较难准确识别恶意样本或家族, 无法挖掘细粒度攻击行为以及定位关键代码. 基于此, 如何构建智能化模型快速、准确地识别海量恶意代码, 分析其家族来源及恶意行为, 有效提取关键特征是 APT 攻击智能检测的关键问题. 该部分将详细综述和分析 APT 攻击检测相关知识, 以整体框架为基础, 详细比较恶意性检测、恶意家族检测、恶意行为识别和恶意代码定位这 4 类关键任务.

2.1 检测框架

当前学术界和产业界关于 APT 攻击检测研究主要根据分析对象分为两大类, 分别是面向终端样本和恶意软件的 APT 攻击智能检测、面向恶意软件生成审计日志的 APT 攻击智能检测. 整个框架如图 4 所示, 前者倾向于产业界通过静态和动态特征提取来检测 APT 攻击, 后者随着溯源图和知识图谱在审计日志分析中应用衍化出对应的模型[12,49].

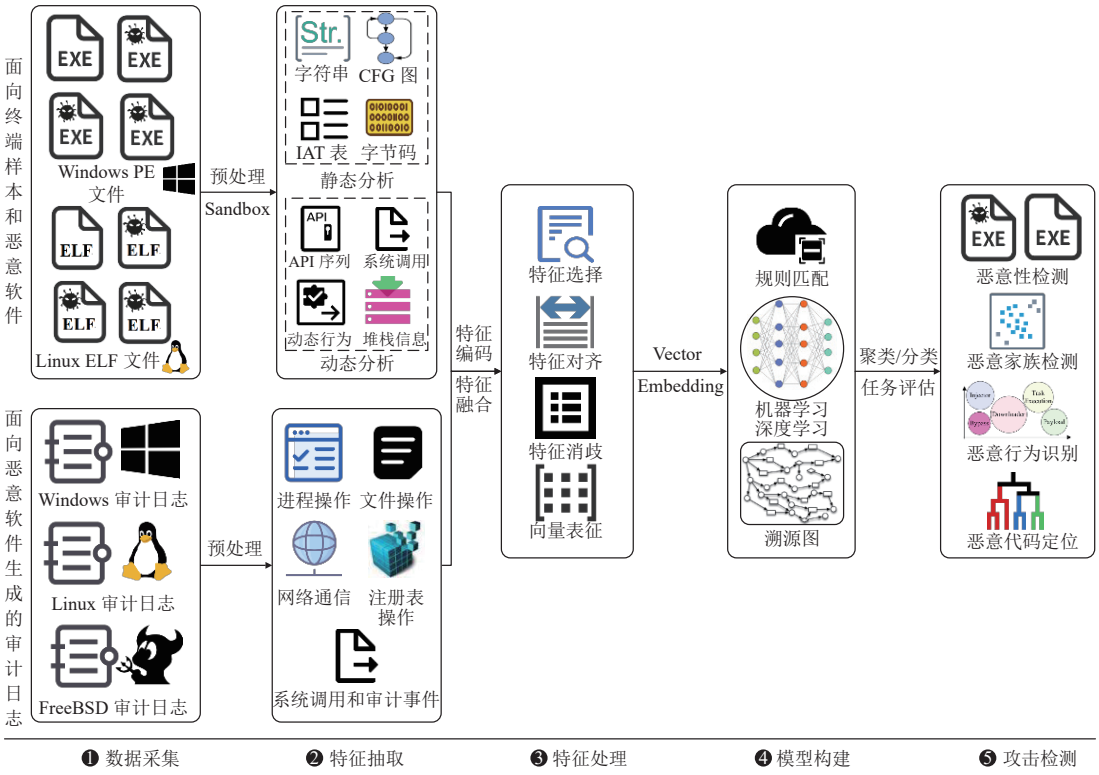

图 4 APT 攻击检测的整体框架

APT 攻击检测的整体框架由 5 部分组成: ① 数据采集包括终端侧各类操作系统的恶意软件, 日志侧集中于威胁情报及流量生成的日志和告警信息; ② 特征抽取旨在利用静态分析工具或动态虚拟环境捕获恶意软件的静态

和动态特征 (如字符串、字节码序列、API 序列等), 利用日志分析系统提取审计日志和告警事件信息 (如进程操作、文件操作、系统调用等); ③ 特征处理旨在利用特征选择、特征对齐、特征消歧来抽取代表性的特征, 并转换成向量或矩阵的形式为后续模型学习; ④ 模型构建包括规则匹配、机器学习、深度学习和溯源图, 通过构建智能模型来实现检测任务; ⑤ 攻击检测旨在利用已构建的模型完成 4 类关键任务, 包括恶意性检测 (含日志侧攻击事件告警识别)、恶意家族检测、恶意行为识别和恶意代码定位.

2.2 现有检测技术综述

该部分将从检测方法、特征类型、溯源关系、知识框架、先验知识、真实场景、鲁棒性和实时检测方面详细比较各类任务的代表性工作, 详细结果如表 2 所示. 需要注意, 由于攻击检测研究相对较多, 并且本文集中于溯源及推理研究, 因此, 该节仅选择代表性工作及方法进行分析, 亦为后续 APT 攻击溯源和推理综述作铺垫.

表 2 APT 攻击智能检测工作总结表

分类	相关工作	检测方法	特征及家族类型	特征分析	溯源关系	知识框架	先验知识	真实场景	DNN	鲁棒性	实时检测
恶意性检测	Liras等人[26]	LR, KNN, RF, SVM	静态特征 (如导入表/语言/可疑API等), 动态特征 (如反调试技术/服务文件等), 网络特征 (IP地址)	动静	×	×	●	√	×	L	×
	AI-HydRa[50]	多层感知机, RF	静态特征 (如文件大小/熵/数量/入口点等), 动态特征 (如网络/API序列等)	动静	×	×	●	√	√	L	×
	HOLMES[51]	溯源图+杀伤链, 高级场景图, 依赖关系剪枝	系统审计日志, 攻击链 (TTP) 和场景知识 (进程/文件/通信/注册表)	动	√	√	●	√	×	M	√
	POIROT[52]	溯源图+查询图, 图匹配+图对齐	系统审计日志, 外部威胁情报知识 (进程/文件/通信/注册表)	动	√	√	●	√	×	M	√
	Foureye[53]	博弈论+蜜罐+策略	网络流量, 自制蜜罐流量特征	动	×	√	●	×	×	L	×
恶意家族检测	Bolton等人[54]	模拟退火算法, 图编辑距离, 随机森林	指令序列特征, 指令路径调用图 [30类家族]	静	×	×	●	√	×	L	×
	微软[55,56]	LSTM, GRU, ESN, Char-CNN, 随机投影算法	指令特征和系统调用序列 [134类家族] (病毒/木马/蠕虫/勒索软件)	动静	×	×	●	√	√	M	√
	Demirkıran等人[57]	Transformer, BERT, LSTM, CANINE	API调用序列及关系 [11类家族] (Trojan/Virus/Worms/后门)	动	√	×	◑	√	√	M	×
	RCNF[58]	二进制灰度图, 胶囊网络, 集成学习	byte和asm文件转灰度图 [9类家族] (Ramnit/Simda/Lollipop等)	静	×	×	●	√	√	M	×
恶意行为识别	Ding等人[59]	CFG-DT, CFG-KNN, CFG-SVM	操作码序列, 控制流程图 [任务: 操作码行为检测]	动静	×	×	●	√	×	M	×
	WATSON[60]	知识图谱, 溯源图, 语义聚合	系统审计日志 (进程/文件/套接字/关系) [任务: APT恶意行为生成]	动	√	√	●	√	×	M	√
	Forecast[61]	知识图谱, 符号分析	内存映像数据和上下文信息, API序列参数 [任务: 功能行为识别]	动	√	√	●	√	×	M	×
	Deep-Reflect[27]	自编码器, HDBSCAN	恶意软件基本块特征, RoI关键区域 [任务: 函数恶意行为识别]	动静	√	×	○	√	√	H	√

表 2 APT 攻击智能检测工作总结表 (续)

分类	相关工作	检测方法	特征及家族类型	特征分析	溯源关系	知识框架	先验知识	真实场景	DNN	鲁棒性	实时检测
恶意代码定位	肖达等人[62]	SVM, RF, KNN, LR, CNN	程序基因和汇编指令 (镜像文件/程序节段/函数/指令流/基本块/汇编指令)	动	×	×	●	√	√	L	×
	SCRUTINIZER[63]	SNN+LSTM, 无监督学习模型, 孪生神经网络	恶意软件重用代码, 运行内存快照, 功能代码片段	动	×	×	◑	√	√	L	×

注: “静”表示采用静态分析方式来提取特征; “动”表示采用动态分析方式来提取特征; ●表示需要大量先验知识或未考虑过度依赖先验知识的问题; ◑表示需要少量先验知识或较少依赖先验知识; ○表示需要极少或不需要先验知识; L (low)表示鲁棒性较差; M (middle)表示鲁棒性中等或具有一定迁移性; H (high)表示鲁棒性好; √表示利用该方法或涉及该内容; ×表示未利用该方法或未涉及该内容

2.2.1 恶意性检测

恶意性检测旨在构建智能模型来学习恶意软件的静态或动态特征, 以及审计日志或网络流量的行为特征, 以判断目标攻击是否具有恶意性. 代表性的方法包括机器学习、深度学习、溯源图和知识图谱. Liras 等人[26]通过提取恶意软件 19 个典型的静态、动态和网络特征, 再构建机器学习模型来识别 APT 攻击. Ghafir 等人[64]提出基于机器学习的 APT 攻击检测系统 MLAPT, 通过威胁检测、警报关联和攻击预测 3 个核心阶段识别 APT 攻击, 并结合 APT 攻击生命周期对网络流量进行实时分析. AI-HydRa[50]是一种融合多层感知机和随机森林的恶意软件检测模型. 然而, 机器学习方法属于浅层学习, 无法捕获深层语义信息且容易被加壳混淆干扰, 较难感知隐蔽性和持续性较高的 APT 攻击.

由于神经网络具有计算速度快、自学习和扩展性更强的优点, 被广泛应用于各个领域, 包括恶意代码分析. 譬如 Mamun 等人[65]构建基于异构任务树的深度学习模型 DeepTaskAPT, 利用 LSTM 网络学习日志中的任务信息和关联关系, 从而识别异常 APT 攻击事件. 图 5 是基于控制流程图 (control flow graph, CFG) 和深度学习的恶意性检测方法, 恶意软件经反汇编和 CFG 图生成提取关键特征, 再转换成对应的向量供模型学习, 最终实现 APT 攻击的恶意性检测, 代表性工作为文献 [66] 和文献 [67]. 此外, 近年来利用溯源图从大规模审计日志中检测 APT 攻击的工作逐渐增加, 通过提取日志告警的主体、客体、系统调用关系及时间戳来构建溯源图, 从而梳理攻击事件并判断其恶意性. HOLMES[51]是一种实时检测 APT 攻击的系统, 旨在将攻击事件的日志特征映射至杀伤链模型, 再利用高级场景图和依赖关系实现 APT 攻击检测. POIROT[52]系统将威胁情报和溯源图关联, 完成对 4 个 APT 组织的恶意性识别. Unicorn[29]系统通过融合概要图和溯源图来识别运行时且长时间潜伏的攻击.

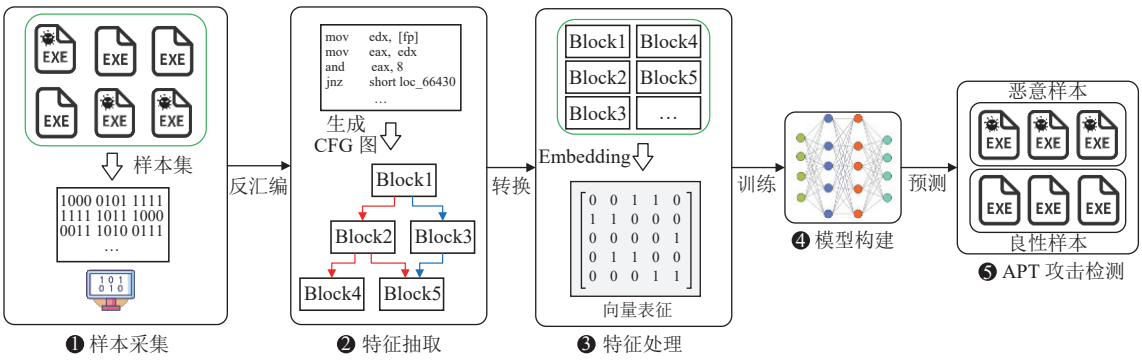

图 5 基于控制流程图和深度学习的恶意性检测方法

总之, 溯源图能捕获攻击事件的细粒度行为, 通过因果关系分析和依赖关联有效刻画攻击场景, 更好地实现 APT 攻击检测. 同样博弈论和数据分析也被用于溯源, 代表性工作为 Foureye[53], 其方法特点如表 2 所示.

2.2.2 恶意家族检测

恶意家族检测是指构建模型识别恶意软件或代码所属家族的类型. 需要注意, 本文将 APT 攻击所属组织的识别划分至溯源任务, 而恶意软件所属病毒、蠕虫、木马或僵尸网络家族的识别定义为恶意家族检测. 典型的方法包括机器学习、深度学习、溯源图以及二进制代码图像化的家族检测方法.

文献 [54] 和文献 [68] 构建机器学习模型 (如随机森林、支持向量机) 来学习恶意软件的静态属性和动态行为, 并预测恶意软件所属家族. 在深度学习方面, 微软公司将随机投影降维算法和神经网络结合, 通过提取大规模二进制文件来实现 134 个恶意软件的识别, 并不断优化其反病毒引擎[55,56]. Gibert 等人[69]通过构建卷积神经网络来学习恶意软件的熵结构特征, 最终实现恶意家族分类研究. 随后, 图神经网络、生成对抗网络和迁移学习逐渐被应用到恶意家族检测领域, 它们能有效学习特征依赖关系, 实现恶意样本增强和减少对先验知识的依赖, 代表性工作为 CFGExplainer[70]、MaliCage[71]和文献 [57].

此外, 将二进制文件转换成灰度图并构建深度学习模型实现恶意家族检测是该任务的代表性方法[58], 不同家族灰度图通常呈现存在一定差异. 然而, 该类方法缺乏足够的解释性并且特征含义丢失严重.

2.2.3 恶意行为识别

恶意行为识别旨在构建模型来预测攻击所属行为或在 APT 攻击生命周期中对应阶段所执行的行为. 通常包括恶意代码基本块或函数的攻击行为研判, 以及 APT 组织常用攻击场景的识别 (如漏洞利用攻击). 代表性的方法包括机器学习、深度学习、溯源图和知识图谱.

Ding 等人[59]设计并实现基于控制流的表示方法, 将 CFG 转换成执行树并生成执行路径, 再提取恶意软件操作码序列作为特征, 并结合机器学习算法实现行为检测, 譬如发现 jmp 跳转指令被用于逃避检测. 文献 [72] 利用有限自动机学习成功识别 APT 攻击常见的 7 种恶意行为 (如病毒复制、端口扫描、shell 命令执行、释放载荷等). 随后, 深度神经网络和溯源图也被用于恶意行为识别研究. MALDC[73]通过 LSTM 模型来学习 API 调用序列构成行为链中的恶意行为. WATSON[60]系统通过基于语义推理和事件聚合来自动提取高级系统行为, 并构建知识图谱有效表征 APT 攻击事件的行为. Forecast[61]系统能够预测恶意软件的功能行为, 覆盖 APT 攻击代码注入、文件执行、资源释放、持久化等阶段. 上述方法丰富和扩展了智能模型在恶意行为识别领域的应用.

此外, 在面向恶意函数或代码块的功能行为识别方面, Downing 等人[27]创新性设计 DeepReflect 系统, 构建无监督的深度神经网络定位恶意软件关键组件或函数的位置, 再利用半监督聚类分析确定恶意函数的行为. Xuan 等人[74]提出一种基于沙箱的恶意软件跟踪系统 Rkprofiler, 能动态监控 Windows 内核恶意软件行为.

2.2.4 恶意代码定位

恶意代码定位是指设计智能模型来提取恶意软件或样本的关键代码或代码基因[32], 以及定位恶意函数的关键代码片段和位置, 它们能在一定程度上丰富 APT 攻击的特征库, 并辅助攻击的检测和溯源.

由于该研究仍处于起步阶段, 相关工作较少. 代表性的工作是肖达等人[62]提出基于程序基因的恶意程序预测方法, 构建机器学习和 CNN 模型来学习所提取的程序基因, 从而实现恶意程序的识别. 李昂[32]设计算法实现对 APT 组织的恶意代码基因提取 (包含可见字符串和汇编代码片段), 有效识别攻击组织的关键特征. SCRUTINIZER[63]系统利用无监督学习模型有效识别 OceanLotus 和 Turla 组织攻击活动中的重用关键代码.

2.3 检测方法

为更好地对比各类智能方法应用于 APT 攻击检测的效果, 本文归纳总结了常见 6 大类方法的特点, 如表 3 所示. 由表可知, 工业界集中于规则特征匹配, 通过安全专家经验来凝练规则和检测攻击; 博弈论具有良好的理论基础, 但其落地应用困难; 机器学习和深度学习是当前常用得方法, 具有良好的智能程度和应用效果; 代码图像化方法缺乏解释性和丢失信息, 较少被工业界和学术界使用; 溯源图和知识图谱方法将是未来研究的重点, 能有效关联攻击行为、还原攻击事件, 并且智能程度较高.

2.4 小 结

本节主要介绍 APT 攻击智能检测研究. 首先给出 APT 攻击智能检测框架; 接着详细综述恶意性检测、恶意

家族检测、恶意行为识别和恶意代码定位 4 个关键任务的相关工作, 并比较现有方法的特点, 为后续 APT 攻击溯源和推理提供帮助. 此外, 随着恶意代码混淆、对抗样本、反调试及反分析技术不断演化, APT 攻击检测存在诸多挑战. 未来期望从智能程度、准确率、鲁棒性、解释性、语义捕获和先验知识等多角度来提升检测效果.

表 3 APT 攻击智能检测工作总结表

方法	优点	缺点	准确率	鲁棒性	解释性	捕获语义	智能程度	检测效率	先验知识
规则特征匹配	适用工业界且快速部署应用	过度依赖专家知识, 耗时耗力	高	低	高	无	低	高	大量
博弈论	具有理论依据和攻防考虑	较难应用于APT攻击防御领域	中	低	低	无	中	中	大量
机器学习	实现简单且容易落地应用	特征工程复杂且依赖专家知识	中	低	低	无	中	高	大量
深度学习	捕获语义知识且提升准确率	复杂度和性能消耗高, 解释性差	高	高	中	高	高	中	中量
代码图像化	能对抗一定混淆	缺乏解释性且丢失特征信息	中	低	无	无	低	中	少量
溯源图/知识图谱	关联行为, 对抗混淆, 解释性好	实现过程复杂, 存在路径爆炸	高	中	高	中	高	低	大量

3 APT 攻击溯源

APT 攻击中的恶意代码会利用混淆、对抗和伪装技术来躲避杀毒软件及防火墙的查杀[5], 甚至会根据攻击目标的实际环境而变化, 导致其攻击痕迹和来源很难被溯源. 因此, 如何快速定位攻击来源, 通过攻击行为和特征知识分析来追踪溯源 APT 攻击的发起组织、区域, 甚至是背后支持的国家、黑客成员和攻击手法至关重要. 它不仅能理解 APT 组织的攻击手法和代码基因特点, 从而制定更智能的防御策略, 还能形成黑客组织、作者和区域的特征库和犯罪证据, 威慑相关黑客组织并保障国家安全[75]. 基于此, 在本节详细总结 APT 攻击溯源的相关知识, 深入挖掘 APT 攻击溯源的特点和原理, 为安全研究人员提供思路并力争从源头遏制 APT 攻击.

3.1 溯源框架

基于现有学术界及产业界 APT 攻击溯源研究的分析与归纳, 本文将 APT 攻击溯源的整体框架划分为 6 个阶段, 包括溯源数据采集、溯源特征抽取、溯源特征处理、溯源模型构建、溯源评分与计算和溯源对象确定, 各阶段间的逻辑关系如图 6 所示, 整个框架以被动溯源模式为主.

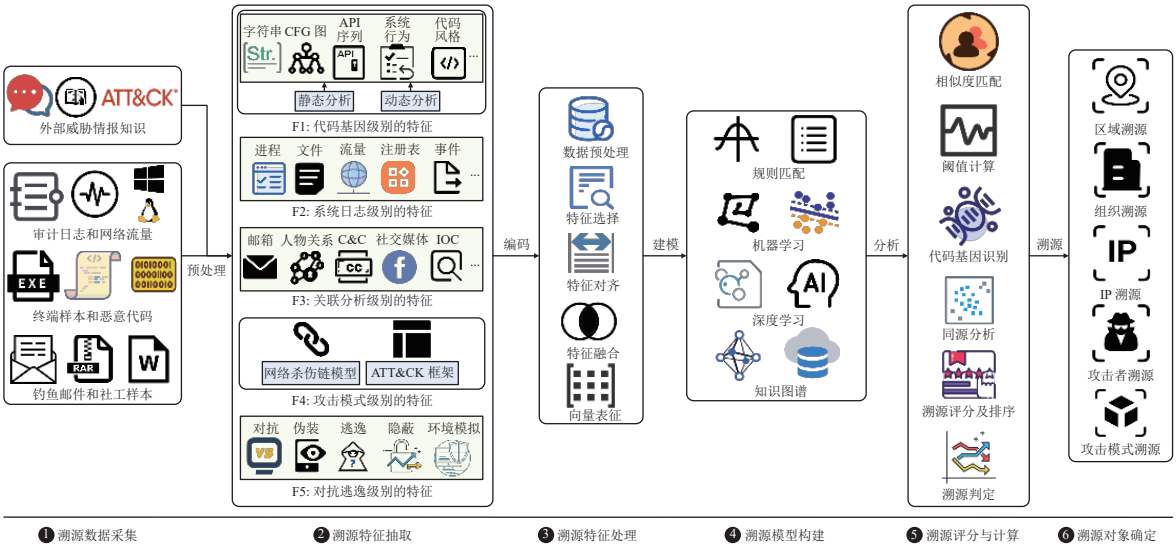

图 6 APT 攻击溯源的整体框架

(1) 溯源数据采集. 在开展 APT 攻击溯源研究前, 通常需要采集大量的样本数据用于分析和训练, 数据集的质量对网络攻击事件追踪溯源至关重要. 本文结合 APT 攻击的数据特性及 TTP 特点, 从流量侧和终端侧将数据源划分为 3 类, 即审计日志和网络流量、终端样本和恶意代码、钓鱼邮件和社工样本, 同时采集外部威胁情报知识 (如 APT 分析报告或 ATT&CK 知识) 来辅助溯源研究. 经过预处理后的数据将流向第 2 个阶段.

(2) 溯源特征抽取. 针对 APT 攻击的隐蔽性、逃逸性和对抗性, 为深层次抽取 APT 攻击溯源特征, 本文归纳了 5 种特征类型, 它们分别是代码基因级别的特征 (如字符串、CFG 图、API 序列等)、系统日志级别的特征 (如进程、文件、注册表等)、关联分析级别的特征 (如邮箱、C&C、社交媒体等)、攻击模式级别的特征 (如网络杀伤链模型和 ATT&CK 框架) 和对抗逃逸级别的特征 (如编译器、系统环境、硬件信息等), 更详细的定义和描述请参考第 3.4 节. 特征抽取是 APT 攻击检测和溯源的重要环节, 通常采用动态分析和静态分析相结合以及自然语言处理等多种方法抽取所需要的特征, 旨在提取具有代表性的特征, 力争有效反映 APT 攻击或恶意代码本质, 这将为后续溯源模型提供丰富的语义知识和特征信息, 从而更准确地溯源目标.

(3) 溯源特征处理. 通常所抽取的原始特征会存在冗余、歧义、不能量化等问题, 溯源特征处理旨在挑选出具有代表性的特征, 并转换为能够被模型识别的矩阵或向量 (常用的向量表征方法为 Word2Vec[76]、Asm2Vec[77]、DeepWalk[78]等), 以提升训练效率并帮助模型更准确地溯源目标. 图 6 中的溯源特征处理包括数据预处理、特征选择、特征对齐、特征融合和向量表征 5 个关键步骤.

(4) 溯源模型构建. 经特征抽取和特征处理后, 接下来将构建溯源模型. 常见方法包括规则匹配、机器学习、深度学习、知识图谱及溯源图. 其中, 规则匹配旨在编写特定规则或正则表达式来匹配溯源特征, 以及构建模型实现后续的相似性计算[79]或同源分析[80]并完成溯源任务, 该方法通常以产业界为主[30,81,82]. 机器学习和深度学习方法旨在构建分类器或聚类模型, 通过对特征向量训练和调参, 直到神经网络参数拟合或模型计算的阈值符合预期条件, 最终对溯源对象进行预测, 典型的方法包括支持向量机 (SVM)[33]、随机森林 (RF)[83]、卷积神经网络 (CNN)[84]、循环神经网络 (RNN)[32]. 基于知识图谱[85]或溯源图[86]的溯源方法是一种新兴方法, 通常会将所抽取的攻击行为特征或审计日志对象 (如进程、文件、socket) 定义为节点或实体, 将攻击事件关联或依赖关系定义为边 (即表示实体之间的不同关系), 最终利用构建的知识图谱或溯源图实施 APT 攻击溯源任务.

(5) 溯源评分与计算. 该阶段是对已构建的溯源模型进行评分、计算和优化, 旨在通过多种操作优化模型并力争匹配到最佳的溯源目标. 其中, 计算操作包括相似度匹配、阈值计算、代码基因识别和同源分析, 结合这些操作和溯源模型能进一步提升溯源效果; 评分操作为溯源评分及排序, 主要是对知识图谱和溯源图的剪枝优化以及多分类模型的实验结果评估, 排序可以遴选出最优的溯源结果; 最终通过溯源判定给出预测目标.

(6) 溯源对象确定. 经过上述 5 个阶段将完成 APT 攻击溯源任务, 本文根据溯源目标对象差异, 将溯源对象划分为 5 大类, 分别是区域溯源、组织溯源、IP 地址溯源、攻击者溯源和攻击模式溯源. 不同目标的溯源方法存在一定的区别, 但其本质相似, 并且可以通过图 6 所示的整体框架完成对 APT 攻击的溯源追踪.

3.2 现有溯源技术综述

基于 5 类溯源对象, 本文从方法、特征、溯源关系图、先验知识、真实场景、鲁棒性和对抗性方面对主流的 APT 攻击溯源方法进行对比分析, 详细比较各种方法之间的优势和缺陷.

3.2.1 区域和国家溯源

Rosenberg 等人[87]首次构建深度神经网络来追踪 APT 攻击的来源, 通过神经网络学习 Cuckoo 沙箱识别的 APT 动态行为来溯源国家信息. 后续该团队又引入迁移学习[88]进一步提升模型的鲁棒性, 更好地溯源国家级 APT 攻击. 安天安全研究与应急处理中心[81]通过逆向分析所提取恶意样本的 C&C、IP 地址和编译器时间戳, 再结合关联分析将白象组织溯源至南亚次大陆地区. Kaspersky 和 Symantec[82,89,90]通过对 Stuxnet 和 Duqu 的功能、结构、代码和数据相似性分析, 发现两者具有相同起源, 被归结为美国和以色列发起的定向攻击. FireEye[91]利用恶意代码分析得到的时区信息、俄语特征和技战术特点, 将 APT28 归因为俄罗斯的黑客组织.

此外, 由于组织、IP 地址及攻击者的溯源一定程度能够识别对应的区域或国家, 因此学术界针对该类溯源研

究的工作相对较少, 而产业界存在过度依赖专家经验和规则库的缺点.

3.2.2 组织溯源

Han 等人[85]提出一种 APT 恶意软件认知框架 APTMalInsight, 旨在提取 APT 样本系统调用信息来描述特征行为, 再构建特征向量和本体知识以溯源 APT 组织. 该方法利用 TF-IDF 算法计算权重, 选取 Top-*N* 的 API 作为特征向量, 最终构建 4 种机器学习方法来完成对比实验. 此外, 该框架能够有效描述攻击行为和理解 APT 攻击的上下文. Bolton 等人[54]提取 APT 样本的子程序指令序列和程序调用图, 再使用模拟退火 (simulate anneal) 算法计算调用图间的编辑距离, 并构建随机森林分类器判别样本所属的组织. 然而, 上述方法未有效考虑 APT 攻击样本的功能特性, 忽视了关键特征对溯源分析的贡献, 并且 APT 组织通常会复用相似功能模块或恶意代码实施攻击[5], 从而导致功能函数之间存在一定的相似性. 基于此, 吕杨琦等人[79]针对 APT 组织复用恶意功能代码现象, 构建 8 类功能的静态检测规则并提取样本的功能函数, 再标准化处理汇编代码以及计算各样本的模糊哈希值, 最终通过各组织特征指纹库匹配识别样本的组织来源. 李昂[32]通过研究如何从恶意代码中提取能够明确指向攻击组织的可见字符串和汇编代码, 并将其定义为该攻击组织的恶意代码基因, 构建基于自注意力机制的双向循环神经网络的汇编代码表示学习模型, 再通过学习向量化的代码基因完成 APT28 和白象组织的溯源实验. Li 等人[92]提出一种基于 SMOTE-RF 算法的 APT 组织分类模型, 该模型将学习经过特征选择的 APT 动态行为数据并有效识别 7 种 APT 组织. FireEye[30]为每个组织生成对应的文档主题摘要 (譬如恶意软件、通信方式), 再利用 TF-IDF 算法生成不同主题的权重, 最后构建相似性度量和聚类算法实现 APT 组织溯源. 赵晶晶[93]提出一种基于决策树的 LightGBM 分类模型, 通过消融和互信息法对多级别恶意软件静态特征 (包括指令级、基本块级、函数级和文件级特征) 进行特征选择, 最终完成组织溯源任务. Laurenza 等人[94]基于安全公司的 APT 分析报告和 VirusTotal 构建恶意软件静态特征知识库, 为每个组织构建一个孤立森林分类器 (isolation forest classifier), 从而预测某个样本是否属于该组织.

然而, 这些方法未有效将静态特征和动态特征结合, 且鲁棒性和可解释性较差. Wang 等人[95]提取二进制文件的静态特征 (代码特征) 和动态特征 (动态行为), 利用 NLP 方法构建特征段落向量和词向量, 再构建随机森林分类器溯源 APT 组织, 并通过 LIME 算法解释分类结果的关键特征. Checkpoint[96]提取俄罗斯两千多个恶意样本的代码及模块特征和 TTP 行为信息, 再识别样本之间的代码相似性, 最终构建关系图谱溯源俄罗斯的 APT 组织, 共计 11 个 APT 组织 (如 APT29、Turla、Sofacy、GreyEnergy 等) 被溯源.

随着溯源图和图神经网络的发展, 基于图论和知识图谱的组织溯源方法被提出. 黄克振等人[33]提出一种基于图模型的网络攻击溯源方法, 通过引入网络攻击事件溯源本体模型来提取攻击事件中的线索数据和威胁情报数据特征, 再经过本体映射和实体对齐生成网络攻击事件溯源关系图, 接着引入图嵌入和随机游走算法学习线索特征并形成事件特征向量, 最终构建 SVM 分类器完成溯源任务. 李腾等人[97]针对日志关系图状态爆炸及缺乏考虑数据隐私保护的问题, 提出一种基于图卷积神经网络 (GCN) 的攻击溯源方法, 利用监督学习解决日志关系状态爆炸问题, 同时优化 Louvain 社区发现算法提升模型的检测速度及准确率, 构建图卷积神经网络实现组织溯源分类, 最后结合加密算法实现对日志数据的隐私保护. 此外, 刘潮歌等人[39]详细总结定向网络攻击的追踪溯源层次化模型, 包括网络服务、主机终端、文件数据、控制信道、行为特征和挖掘分析 6 个层次, 并从纵横体系和多维度追踪溯源定向网络攻击, 以及对白象 APT 组织溯源取证分析. Lemay 等人[37]系统总结了全球 40 个 APT 组织的特征及技战术特点. 宋文纳等人[14]从学术界和产业界系统梳理了恶意代码溯源与演化技术.

总之, 当前针对 APT 组织的溯源研究相对较多, 而攻击者的溯源研究较少. 如何细粒度识别恶意代码作者和攻击事件的发起者至关重要, 将有助于积累犯罪证据威慑对应的黑客成员.

3.2.3 攻击者溯源

通常分为源代码作者溯源和二进制代码作者溯源. 在源代码作者溯源方面, Krsul 等人[98]构建机器学习模型来检测编程风格特征以识别程序作者, 其中编程习惯定义为编程布局、编程风格和编程结构这 3 类. Burrows 等人[99]将信息检索技术应用于 C 语言代码作者的归因问题, 通过定义编程风格 (如注释、命名规则、排版布局)、控制结构样式 (如操作符、关键字、标准库影响算法决策的控制令牌等)、信息结构风格 (如程序内存、数据结构、输入

输出函数等代码结构信息) 等 3 类特征, 再筛选 6 种关键特征 (即空白符、运算符、数据类型、关键字、I/O 关键字、标准库函数) 并应用于信息检索系统以完成作者溯源. 然而, 上述方法忽略代码结构习惯的重要性, 并且较难被伪装和克隆. 因此, Chen 等人[100]提出一种基于程序依赖图 (program dependence graph, PDG) 的源代码作者溯源方法, 通过计算程序内部数据流的相似性来识别作者身份. Caliskan-Islam 等人[101]针对代码风格、代码文本特征和语法特性, 提出一种基于模糊抽象语法树 (abstract syntax tree, AST) 和随机森林算法的匿名作者识别方法. 随后, Dauber 等人[102]在源代码构建抽象语法树的基础上, 引入系统 API、关键词和 Unigram 特征并转换为特征向量, 再建立随机森林模型实现作者溯源, 并在开源代码协作问题上完成多作者识别任务. Alsulami 等人[103]将神经网络作用于抽象语法树, 利用 LSTM 模型自动学习 AST 表示的特征, 从而实现作者溯源. Abuhamad 等人[104]提出一种基于深度学习的代码作者身份识别系统 (DL-CAIS), 该系统通过 TF-IDF 表征向量, 利用循环神经网络和随机森林算法实现作者溯源, 并能有效对抗混淆且识别 4 种编程语言 (C、C++、Java 和 Python) 的作者.

然而, 编程风格和代码布局通常会受限于源代码, 经过编译后的二进制代码会丢失源代码的风格特征 (如注释、布局、命名), 从而导致先前源代码作者溯源的部分方法对二进制代码的作者溯源不再适用. 此外, 恶意软件和 APT 攻击样本多数缺少源代码且以二进制文件为主, 先前方法无法将功能特征与作者风格准确关联, 并且在网络安全领域应用较少. 因此, 如何提取恶意二进制文件的关键代码和风格特征是攻击者溯源的关键.

基于此, Rosenblum 等人[105]首次针对二进制代码提出一种自动检测代码风格特征的方法, 该方法能提取代码的内部结构、控制流图 (CFG)、代码风格和函数调用特征, 利用聚类算法确定程序的作者以及发现未知程序作者风格的相似性. 在 Blackhat 会议上, Marquis-Boire 等人[106]根据人工经验定义 4 类特征, 基于恶意二进制文件的可信链接来识别相同的作者. 其中, 4 类特征分别是字符串特征 (如时间戳、C&C 地址)、实现策略特征 (如内存分配习惯、动态 API 序列、编译器信息)、自定义特征 (如加密压缩算法、传播机制、混淆技术)、基础设施特征 (如国家语言、用户代理、通信协议及端口). 尽管该工作是早期针对 APT 攻击的研究, 但存在过度依赖人工经验且需要大量手工处理提取特征的缺陷. 随后, Alrabaee 团队[31,84,107]提出多种自动识别二进制文件作者身份的方法. 在 2014 年, 他们提出一种多层次作者溯源方法 OBA2[107], 该方法首先在阻塞层通过签名方法检测与作者风格无关的功能; 接着在代码分析层建立二进制代码语法和已标识源代码的映射, 为每个作者创建概要, 在寄存器流分析层构建寄存器流图来提取二进制文件的语义特征; 最后通过规则匹配溯源作者并在 GCJ (Google code jam) 数据集上实验. 2016 年, Alrabaee 等人[31]详细分析和总结源代码文件及二进制文件作者溯源研究的常见特征, 并且提取编译影响较小且与作者编程风格关联性更高的特征作为关键特征, 接着转换为特征向量并计算特征与作者间的互信息, 再利用排序算法和 SVM 分类器完成恶意软件作者的溯源. 在 2019 年, 该团队[84]又提出 BinEye, 它是一种基于深度学习的二进制作者溯源工具, 由 3 种卷积神经网络组成, 所使用的特征包括二进制文件转换的灰度图、可执行文件转换的字节码和用户函数的操作码. 该工具能在较少先验知识的情况下自动学习每位作者的编程风格, 并且表征程序二进制文件作者以完成溯源任务. 乔延臣等人[108]提出一种基于 WinAPI 调用习惯的恶意代码自动同源判定方法, 该方法首先定义 7 类 WinAPI 调用行为, 再使用数据挖掘算法构建作者编程习惯模型, 利用频繁模式利群因子计算样本的同源度, 再通过 K 均值聚类算法预测同源判定阈值并完成恶意代码作者的同源分析. Caliskan 等人[83]通过反汇编提取指令、符号和字符串, 利用反编译器提取语法和控制流特征, 再将代表性特征与抽象语法树关联, 训练一个随机森林分类器实现二进制代码的作者识别, 并在真实的“在野”代码及黑客论坛程序中实现作者去匿名化. Alrabaee 等人[109]提出一种新型方法来识别二进制程序的作者. 该方法称为 CPA, 通过作者代码习惯、代码结构特征以及作者实际编程业务知识来构建作者风格特征, 利用 LDA (latent dirichlet allocation) 算法生成作者风格签名, 从而帮助识别具有相似风格二进制文件的作者, 最终实现跨平台二进制代码的作者识别. 针对软件代码会因变形、混淆或伪装导致同源特征难以表征, 吴鹏[110]提出一种基于混合语义的二进制代码同源判定方法, 通过融合文本语义和代码结构语义, 实现文本嵌入和图嵌入的混合语义表达, 利用孪生神经网络模型实现二进制代码的同源分析. 此外, 由于网络攻击会产生大量的日志和流量信息, 因此针对审计日志的作者溯源研究产生. 王梓晗[111]根据勒索软件攻击防御的特性, 提出一种勒索软件的追踪溯源模型, 利用网络欺骗技术诱导攻击并采集网络日志及终端数据, 再结合机器学习和自然语言处理提取攻击者线索并实现勒索软件的作者溯源.

Berady 等人[112]从攻击者和防御者的角度剖析攻击元素, 结合 TTP 技术和共同的对象及组件来抽取攻击痕迹并构建溯源图, 再不断调整规则和模型来提升攻击者检测效率, 最终以 APT29 的模拟攻击事件进行实验.

尽管基于源代码和二进制代码的作者溯源取得了一定成果, 但这些方法仍需要借助大量的预处理和专家知识来提取特征, 并且源代码作者的编程风格和代码习惯容易被篡改和伪装, 经过编译或混淆会丢失关键特征信息, 而二进制代码的关键特征及代码风格较难被提取, 且缺少有效解决二进制代码和源代码映射的语义鸿沟问题. 此外, 当前研究主要集中于恶意代码家族分类, 针对恶意代码作者的溯源研究仍处于发展阶段, 自动化和智能化的攻击者溯源研究亟需加强.

3.2.4 IP 或 C&C 地址溯源

在真实的网络攻击中, 黑客组织会利用源地址伪造、IP 跳板/代理等技术伪装、模拟或隐藏攻击, 导致攻击组织和成员较难溯源. 此外, APT 组织通常会开展跨国家跨区域的网络攻击, 尤其是利用僵尸网络的 DDoS 攻击, 这会造成 IP 数据流分类困难并给攻击源追溯带来巨大挑战[113]. 因此, 如何准确识别攻击的 IP 地址或 C&C 地址变得至关重要, 它将有助于追踪攻击流量的来源和身份.

在 IP 地址溯源方面, Tian 等人[114]提出基于 IP 回溯采样流 (SampleTrace) 的增量部署方法, 从理论上分析通过 IP 流追踪不同 AS 跳数的概率以及攻击流包数之间的定量关系. 徐恪等人[115]从研究体系、实现机制和关键技术 3 个维度对地址安全技术进行归纳, 对它们的地址欺骗防御能力、可部署性和开销等指标进行评估. 实践层面, 针对攻击源地址容易被伪造且较难被准确检测的问题, Yu 等人[116]提出一种基于确定包标记溯源 (determinic packet marking, DPM) 机制的按需标记 (marking on demand, MOD) 溯源方案, 再通过 DPM 策略标记路由器, 当监视器在网络流量中发现可疑目标时, 会标记相关请求的 IP 地址, 受害者再从攻击包中提取标记指向 MOD 服务器的请求, 利用映射关系来获取攻击源并识别 DDoS 攻击. 然而, 面对大规模攻击时, 现有方法存在溯源负载受限、扩展性差的问题. 因此, 鲁宁等人[113,117]提出一种可扩展的动态确定包标记溯源方法 (SEEK), 借鉴 IP 地址分层复用思想划分标签空间, 利用层次化的溯源联盟体系结构和动态调整包标记概率来增加溯源规模并均衡设备负载, 再通过动态扩展标签装载域和自适应管理标签策略来提高标签的可用量和利用率, 最终增强 IP 溯源系统的扩展性. 姜建国等人[118]将网络攻击源追踪技术分为虚假 IP 追踪、Botnet 追踪、匿名网络追踪、跳板追踪和局域网追踪这 5 类问题, 并进行详细综述和总结, 包括对 IP 地址和 C&C 地址的溯源研究.

在 C&C 地址溯源方面, Kaspersky 公司[82]利用关联分析和同源分析技术发现 Gauss 和 Flame 恶意软件在代码、架构和 C&C 通信上存在较多的相似特征, 并有效溯源其攻击地址, 最终证实它们是“亲属关系”. Hong 等人[119]针对 APT 攻击通常会使用多个命令和控制 (C&C) 通道来躲避检测, 提出了 Ctracer 系统, 整个系统从网络流量和审计日志中识别 C&C 会话的共享网络签名来溯源攻击, 并在企业环境中利用 C&C 成功识别出 APT 攻击. Oprea 等人[120]利用基于图论启发的信念传播框架来检测 APT 攻击中的早期感染问题, 并构建一个针对企业环境的 C&C 通信检测器以追踪其他相关域的可疑主机或活动. Fuller 等人[121]研究发现过度授权协议提供一种秘密监视 C&C 服务器的方法, 他们采集 2006–2021 年的 20 万个恶意软件, 发现有 62 202 个机器人和 443 905 个 C&C 监控功能存在协议过度授权的问题, 这也是僵尸网络中断和宕机的根本原因. 基于此, 他们提出 C3PO, 一种自动识别过度授权协议的管道, 通过基于内存图像的符号执行和 API 功能映射有效识别僵尸网络和后渗透通信中的过度授权现象, 并发现渗透向量和 C&C 监控能力, 从而有效促进 C&C 地址的溯源工作.

然而, 尽管针对 IP 地址和 C&C 地址的溯源研究已取得一定进展, 但由于 APT 攻击潜伏时间较长, 特定组织通常会利用逃逸样本、变种木马、暗网通信, 以及组建僵尸网络、构造无文件攻击等手段瓦解内网安全, 并且这些网络攻击中的 IP 地址或 C&C 地址会不断跳转及演化, 甚至伪装成其他目标来干扰安全工作者的溯源研究. 因此, 现有方法仍然无法精准追踪 APT 攻击各个阶段的目标地址, 这也是整个安全界未来将深入研究的问题.

3.2.5 攻击模式溯源

根据 360 公司[9]最新溯源研究发现, 当前国家级 APT 组织通常会使用多个跳板机和 0day 漏洞来躲避检测和溯源, 譬如美国国家安全局 (NSA) 针对某大学的网络攻击, 先后精心挑选 49 台跳板机, 同时结合多个 0day 漏洞 (如 EXTREMEPARR 和 EBBISLAND) 发起攻击并窃取情报. 同时, 根据 Checkpoint 公司[122]调查发现, 大型黑客

组织通常需要不同的团队协作来完成网络攻击, 各团队需要调用不同模块或 API 组件来构建恶意程序, 甚至会模拟或使用其他组织的攻击代码. 因此, 先前的溯源方法较难定位到特定的组织或成员, 基于攻击模式的溯源工作从另一个角度提供线索, 挖掘网络攻击细粒度和抽象层面的模式并锁定 APT 攻击的潜在范围 (即区域或组织), 从而辅助 APT 攻击溯源.

然而, 当前缺乏系统性分析攻击模式的溯源工作, 大多数研究是结合 ATT&CK 框架来刻画 APT 组织的攻击模式或攻击画像, 再利用实际攻击产生的日志或流量信息来构建溯源图或因果关系图, 最终实现网络攻击检测和溯源. 比较具有代表性的工作是 ProTracer 系统[86]、TRACE 系统[123]以及定向网络攻击追踪溯源层次化模型[39]. 其中, TRACE 溯源跟踪系统通过静态分析来识别程序单元结构和单元间的依赖关系, 再结合攻击模式构建因果关系图并识别 APT 攻击. ProTracer 是一种轻量级的溯源跟踪系统, 它通过将程序划分为多个单元, 再利用 Tracepoint 实现用户空间进程和内核代码间的细粒度污点跟踪, 对后门攻击、信息窃取、非法存储和钓鱼邮件 4 个场景实现因果关系图的重建, 从而完成攻击者的调查取证和溯源定位.

综上所述, 本文按溯源目标从区域、组织、地址、攻击者和攻击模式 5 个视角进行研究, 并归纳分析现有工作在相关领域的应用和已取得的成果. 然而, 当前仍缺少对 APT 攻击溯源研究进行系统、深入、全面地梳理和总结的工作, 也缺乏智能化和自动化的溯源研究. 基于此, 本文从学术界和工业界对 APT 攻击智能溯源领域的现有方法及最新研究进展进行归类和综述, 并对比不同阶段、不同方法的优缺点. 表 4 详细比较现有 APT 溯源工作. 其中, “溯源方法”“溯源特征”和“溯源目标”描述该研究的解决方案; “攻击类型”和“数据源”旨在区分真实网络攻击、APT 攻击和特定的攻击场景, 并表明智能攻击的类型和实验数据; “特征分析”是特征提取所采用的方法, 包括静态分析、动态分析和动静态结合分析; 其他指标旨在从多个维度评估 APT 溯源方法, 包括智能性、真实性、鲁棒性和对抗性等.

以上相关工作对比见表 4. 整体而言, 现有工作仍没有较好地解决 APT 攻击追踪溯源问题, 它们均依赖大量的先验知识和专家经验, 尤其在特征抽取和特征处理阶段 (详见图 6 的溯源框架). 企业界更倾向于利用规则匹配和同源分析实现区域、组织和攻击者的溯源; 学术界更倾向于利用机器学习、深度学习或自然语言处理技术构建模型实现溯源追踪. 当前该领域利用溯源图和知识图谱挖掘潜关联关系并识别攻击目标的工作较少, 且鲁棒性和迁移扩展性较差, 也缺乏对溯源对抗、伪装、模拟和针对攻击模式层面的研究.

表 4　APT 攻击智能溯源工作总结表

相关工作	溯源方法	溯源对象	溯源特征	攻击类型	数据源	特征分析	溯源关系图	实验验证	先验知识	真实样本	DNN	鲁棒性	对抗性
DeepAPT[87]	DNN	区域 国家	动态行为, 静态特征	APT	自定义数据集 (Cuckoo沙箱)	动静	×	√	●	√	√	L	×
Rosenberg[88]	DNN, 迁移学习	国家	动态行为	APT	自定义数据集 (Cuckoo沙箱)	动	×	√	◑	√	×	M	×
安天安全研究与应急处理中心[81]	规则匹配, 逆向分析, 关联分析	区域 攻击者 C&C	时间戳, 代码和行为特征, 社交媒体特征	APT	企业数据集	动静	×	√	●	√	×	L	×
黄克振等人[33]	图模型和本体, 溯源关系图+SVM	组织	攻击事件线索, 安全威胁情报	APT	模拟数据 FireEye	动静	√	√	◑	√	×	L	×
李昂[32]	恶意代码基因, Instruction2Vec, RNN	组织	恶意代码字符串, 汇编代码	APT	自定义数据集, APT malware, VirusTotal	静	×	√	●	√	√	M	×
吕杨琦等人[79]	规则匹配, 模糊哈希	组织	恶意代码, 功能函数	APT	自定义数据集, AnyRun	静	×	√	●	√	×	L	×
李腾等人[97]	图卷积神经网络, Louvain算法	组织	日志及流量数据	APT	CCCS-CIC- AndMal-2020 仿真流量数据	动	√	√	●	×	√	L	×
Han等人[85]	本体和知识图谱, RF DT KNN XGB	组织	动态行为和调用, API序列	APT	自定义数据集, VirusShare	动	√	√	●	√	×	L	×

表 4 APT 攻击智能溯源工作总结表 (续)

相关工作	溯源方法	溯源对象	溯源特征	攻击类型	数据源	特征分析	溯源关系图	实验验证	先验知识	真实样本	DNN	鲁棒性	对抗性
刘潮歌等人[39]	定向网络攻击, 追踪溯源模型	组织	归纳网络和系统环境的6类特征	APT	—	—	×	×	●	×	×	L	×
赵晶晶[93]	LightGBM	组织	代码特征和风格, 基本块和指令	APT	自定义数据集	静	×	√	●	√	×	L	×
FireEye[30]	TF-IDF+聚类, 相似性计算	组织	组织统计信息, 文档主题特征	APT	企业数据集	动	×	√	●	√	×	×	×
APT ecosystem[96]	关系图谱+聚类, 代码同源分析	组织	动态行为, 代码和TTP信息	APT	企业数据集	动静	√	√	●	√	×	×	×
Wang等人[95]	NLP RF+DNN	组织	代码函数特征, 动态行为特征	APT	自定义数据集, VirusTotal	动静	×	√	●	√	√	L	×
Alrabaee等人[84,107]	规则匹配+相似性, 机器学习	攻击者	二进制代码特征, 编程风格特征	恶意代码	GCJ数据集, 自定义数据集	动静	×	√	●	√	×	L	×
王梓晗[111]	网络欺骗环境, 自然语言处理	攻击者	审计日志, 溯源线索	勒索软件	模拟数据	动	×	√	●	×	×	L	×
Checkpoint[124]	关联分析, 逆向分析	攻击者 C&C	代码和流量特征, 社交媒体特征	APT	企业数据集	动静	×	√	●	√	×	L	×
Caliskan等人[83], Dauber等人[102]	RF+SVM	作者	代码特征, 指令和CFG图	开源代码	GCJ数据集	静	×	√	●	×	×	L	×
Berady等人[112]	关系图谱, 规则匹配	攻击者	审计日志, IOC信息	APT	模拟数据	动	√	√	●	×	×	L	×
鲁宁等人[113, 117]	联盟模式, TIST	IP	IP流和链路指纹	网络攻击	仿真拓扑数据	动	×	√	●	×	×	L	×
C3PO[121]	符号执行(iSSE), API功能映射	C&C 监控	服务器监控事件, API调用和CFG	僵尸网络	自定义数据集	动	×	√	●	×	×	L	×
Kaspersky[82]	关联分析, 同源分析	C&C	代码函数特征, 动态行为特征	APT	企业数据集 Flame+Gauss	动静	×	√	●	√	×	L	×
Ctracer[119]	聚类	C&C	审计日志, 网络流量	APT	流量模拟数据	动	×	√	●	×	×	L	×
ProTracer[86]	溯源图, 内核代码跟踪	攻击模式	审计日志	网络攻击	自定义数据集	动	√	√	●	√	×	M	×
TRACE[123]	因果关系图	攻击模式	程序单元结构, 单元依赖关系	APT	模拟数据	静	√	√	●	×	×	L	×

注: “静”表示采用静态分析方式来提取特征; “动”表示采用动态分析方式来提取特征; ●表示需要大量先验知识或未考虑过度依赖先验知识的问题; ◑表示需要少量先验知识或较少依赖先验知识; ○表示需要极少或不需要先验知识; L (low)表示鲁棒性较差; M (middle)表示鲁棒性中等或具有一定迁移性; H (high)表示鲁棒性好; √表示利用该方法或涉及该内容; ×表示未利用该方法或未涉及该内容

3.3 溯源机理

随着信息技术飞速发展, 高级可持续性威胁 (APT) 攻击事件频发, 给全球政治、经济、科技、教育、文化和网络空间造成极其严重的影响. 如何有效溯源攻击事件的源头, 定位攻击者所属区域和组织, 挖掘攻击者的真实身份和攻击路径, 并有效实施法律追责和网络防御已成为重要的研究问题. 近年来, 学术界和工业界的科研人员利用恶意代码分析、数字取证、蜜罐、大数据分析、威胁情报关联、人工智能等技术开展网络攻击追踪溯源研究[33,125], 溯源技术被广泛应用到多个领域 (如网络犯罪取证和网络攻击追踪), 俨然已经发展为一项新兴技术. 接下来本文将给出“网络攻击溯源”和“APT 攻击智能溯源”的定义, 再介绍现有溯源方法分类.

(1) 溯源定义

溯源在英文中通常表示为“Attribution”“Traceback”或“Tracking”, 旨在追踪和定位网络攻击源头. 在汉语中, 溯

源是指寻找发源之地, 用来探寻事物的根本和源头, 已被广泛应用于犯罪取证、医学生物、文化考古等领域. 本文主要针对 APT 攻击开展溯源综述研究. 首先, 结合现有工作的 7 种网络攻击溯源定义[5,14,33,40,80,118,126], 给出 APT 攻击溯源的一般性描述及七元组形式化定义.

分析发现, 传统攻击溯源以确定攻击者身份和位置为主[126], 主要针对网络攻击. 工业界[80]倾向于借助专家经验, 利用同源分析或关联分析技术聚合攻击事件并形成知识库, 或精心构建蜜罐和欺骗系统, 再从中寻找痕迹和关联以实现对网络攻击的取证及溯源, 最终保护内部资产和系统安全, 比较典型的应用如安天公司的捕风蜜罐系统和追影威胁分析系统[127]、奇安信公司的网神攻击诱捕系统[128]、FireEye 和 McAfee 公司的 Trellix XDR 平台[129]. 学术界倾向于通过提取攻击或样本的特征, 利用规则匹配、大数据分析、机器学习、深度学习等技术构建模型, 从而预测溯源攻击者相关信息, 包括组织、地址、作者身份等.

然而, 先前的研究较少考虑 APT 攻击、高隐蔽攻击和定向网络攻击的复杂性、隐蔽性、持续性和对抗性, 溯源对象囊括不全, 定义存在一定的局限性, 无法有效覆盖日益变化的网络攻击. 此外, APT 攻击的发起者具有极强的组织性和专业性, 通常会利用先进的攻击技术 (如 0day 漏洞[130]或无文件攻击[44]) 寻找目标网络的薄弱点, 黑客组织甚至会组装包含一系列正常行为的定制恶意样本发起攻击, 也可能会利用混淆或伪装技术模拟其他组织的攻击, 从而绕过现有防御技术和溯源系统. 综上, 为了有效推动 APT 攻击智能溯源研究的发展, 更准确地将虚拟地址和现实世界的区域位置、攻击互联网中的账号和真实物理空间中的身份及组织关系关联, 本文开展 APT 攻击溯源的详细综述和分析, 并给出广义的一般性描述和形式化定义.

定义 1. APT 攻击智能溯源. 针对 APT 攻击或定向网络攻击各阶段特点, 利用关联分析、机器学习、深度学习、知识图谱等智能化技术提取溯源特征并构建溯源模型, 以学习和理解攻击活动 (含环境和对抗信息) 中的语义知识, 并深入挖掘攻击事件中的线索和痕迹, 发现虚拟网络与现实物理空间的对照关系, 关联国家组织成员代码间的衍生关系, 最终实现对攻击源头信息的追踪和预测, 溯源对象覆盖攻击区域、组织、地址、作者身份和攻击模式. 整个 APT 攻击智能溯源的形式化定义可以用公式 (1) 表示.

$$APTAttackAttribution = \langle AT, AE, AF, AM, AC, AP, AO \rangle \tag{1}$$

其中, AT 表示溯源目标 (attribution target), 旨在寻求发起攻击的缘由和目的, 如 Stuxnet 是针对伊朗核工业基础设施的网络攻击[89,90]; AE 表示溯源环境 (attribution environment), 包括实施攻击硬件环境 (如系统版本)、软件环境 (如 Office 版本)、网络环境 (如端口信息) 和攻击场景信息, 其中攻击场景是指 APT 组织常用攻击手法所实施的目标场景, 典型的譬如鱼叉式钓鱼网络攻击、恶意代码攻击、漏洞利用攻击、水坑攻击等; AF 表示溯源特征 (attribution feature), 本文是指从攻击活动中所提取用来追踪攻击目标的特征, 如进程、系统行为、文件、流量、注册表等, 本文归纳总结了 5 种级别的特征, 详见第 3.4 节; AM 表示溯源模型 (attribution model), 是利用关联分析、机器学习、深度学习和知识图谱所构建的智能溯源模型, 能有效学习溯源特征潜在的规律和关联; AC 表示溯源对抗 (attribution confrontation), 在溯源过程中 APT 攻击会利用混淆、伪装、欺骗、模拟等多种手段来对抗溯源和逃逸检测, 因此, 为更好地实现智能溯源、抵抗反溯源机制, 本文将溯源对抗也归纳在形式化定义中; AP 表示溯源过程 (attribution process), 具体是指追踪攻击完整生命周期 (如图 1 所示) 的过程, 从溯源发起到溯源结束, 涵盖攻击的各个阶段; AO 表示溯源对象 (attribution object), 是指完成整个溯源任务的目标对象, 包括攻击者的虚拟身份信息和物理身份信息以及与其相关的关联信息, 本文经过详细地综述和分析, 将溯源对象划分为 5 类, 分别是攻击区域、组织、地址、作者身份和攻击模式, 其中区域和攻击模式属于粗粒度对象, 组织、地址和作者身份属于细粒度对象.

综上, 本文给出了 APT 攻击智能溯源的精确定义和描述, 旨在更好地帮助学术界和工业界快速定位和追踪攻击来源, 实现智能化和自动化溯源研究. 本文提出的溯源定义及框架能覆盖多种类型、不同场景的攻击, 涵盖 APT 攻击、定向网络攻击、网络攻击和恶意代码攻击, 不同攻击场景均能从图 6 的溯源框架中匹配到一条潜在的智能溯源路线, 即使面对溯源对抗, 也能有效实现攻击的溯源和追踪.

(2) 溯源分类

先前研究中, Cohen 等人[131]将网络攻击溯源划分为攻击主机、控制主机、攻击参与人和攻击组织 4 个级别; 刘潮歌等人[39]将溯源划分为主动溯源和被动溯源; 宋文纳等人[14]按照学术界和企业界进行分类; 陈瑞东等人[5]将 APT 防御方案分为 4 类, 包括恶意代码检测类、网络入侵检测类、主机应用保护类和大数据分析检测类; 姜建国

等人[118]针对网络攻击真实地址通常会被隐藏, 按照问题描述将攻击源追踪划分为虚假 IP 追踪、匿名网络追踪、Botnet 追踪、跳板追踪和局域网追踪 5 大类; 工业界[90,128,129]通常按照主机终端侧和网络流量侧构造溯源系统.

本文主要结合溯源目的和 APT 攻击生命周期各阶段流程, 按照溯源对象进行分类, 最终划分为区域溯源、组织溯源、地址溯源、作者身份溯源和攻击模式溯源 5 大类. 介绍完溯源机理后, 本文将归纳和总结常见溯源特征和溯源方法.

3.4 溯源特征与方法

本节将详细介绍常见的溯源特征和溯源方法.

3.4.1 溯源特征

溯源是融合软件逆向技术、数字取证、人工智能、社会工程和网络攻防等技术以实现攻击源追踪的一种高级防御手段. 针对 APT 攻击溯源研究, 安全厂商通常会利用网络威胁情报分析实现追踪溯源, 典型的 IOCs (indicator of compromises, 威胁标记) 特征包括恶意文件的 HASH 值、主机特征 (如互斥体、运行路径、注册表项)、网络特征 (IP、域名、URL、通信协议)、事件特征、组织特征以及人员情报等, 如图 7 所示, 金字塔越往上其溯源难度越大. 与之对应, 学术界通常采用静态特征、动态特征和动静态相结合的特征开展网络攻击溯源研究[14,49]. 溯源特征是 APT 攻击智能溯源的重要支撑, 它决定了所构建模型是否能从网络流量或审计日志中学到区分区域、组织或攻击模型的本质, 也决定了是否能准确发现攻击源地址和攻击者身份. 不同于先前的工作, 本文针对 APT 攻击的隐蔽性、逃逸性和对抗性, 结合 APT 攻击溯源的整体框架 (图 6) 和威胁情报溯源的层次梯度金字塔 (图 7)[132], 深入挖掘 APT 攻击的溯源特征, 并归纳总结 5 大类特征, 如表 5 所示.

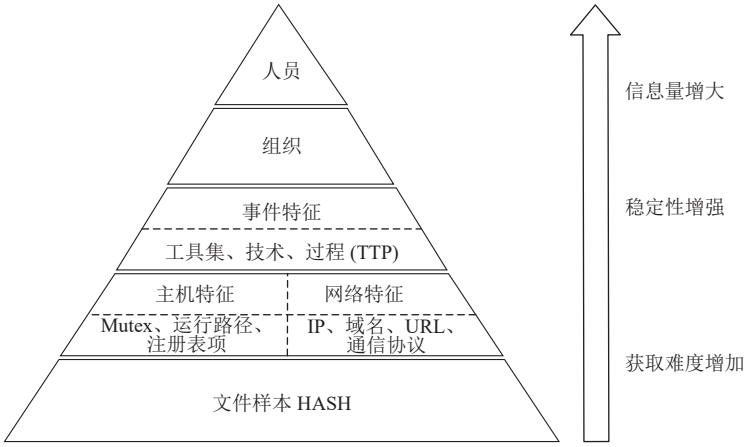

图 7 威胁情报溯源的层次梯度金字塔

表 5 APT 溯源特征归纳与总结

类别	含义	特征类型	相关文献	特征优点	特征缺点
代码基因级别特征	是指能够有效标识恶意软件或样本, 携带独特的功能代码片段或承载关键行为信息的特征, 类似于生物遗传学中的基因[32]	① 静态特征 字节码序列、字符串、编码风格、PE文件头部信息 (含导入表、导出表、动态链接库、节信息、时间戳)、控制流图 (CFG)、程序依赖图 (PDG)、抽象语法树 (AST)、IOC、灰度图、编译信息、语种、证书、权限、熵	[31, 32, 54, 79, 83, 87, 93, 94, 95, 96, 99, 100, 101, 102, 105, 109, 121, 123]	部分特征易获取、特征提取速度快、特征种类丰富、能直观反映样本	易被加壳加密混淆干扰、需要人工标注、缺乏语义信息易被针对、难以反映行为
		② 动态特征 API序列、系统调用行为 (含网络通信行为、文件行为、注册表行为)、运行时状态信息 (含寄存器、CPU、内存、线程、IO)、API-CFG	[80, 85, 87, 88, 92, 95, 96, 108, 121]	包含丰富语义信息、能反映样本行为、能规避混淆对抗	动态分析系统开销较大、遇到反动态分析时特征提取难

表 5 APT 溯源特征归纳与总结 (续)

类别	含义	特征类型	相关文献	特征优点	特征缺点
系统日志级别特征	是指由审计系统或配置服务器按时间顺序记录日志事件和告警所形成的特征, 旨在快速发现潜在的攻击事件, 对各类风险和威胁执行实时监控[133]	进程 (线程)、文件、Socket网络通信 (IP地址+端口号)、注册表、审计日志事件	[33, 86, 97, 111, 112, 119, 120, 121, 123]	有效记录系统日志事件、易与溯源图结合追踪攻击源、包含因果依赖关系和时间顺序、能辅助运行时检测	易存在状态依赖爆炸问题、较难区分正常和恶意事件、特征粒度较粗很难发现溯源痕迹
关联分析级别特征	是指利用关联分析和社会工程学挖掘APT攻击来源及作者身份信息的特征集合, 安全公司通常利用该级别的特征溯源APT攻击[30]	① 社会工程学关联特征 邮箱、电话、社交媒体账号 (如Twitter、GitHub、Facebook)、出生日期、人物形象、含经纬度照片、履历 ② 样本IOC关联特征 C&C地址、IP地址、域名、域名注册邮箱、载荷HASH值、图标信息、URL、Mutex、PDB、启动路径、软件信息、硬件信息 ③ 组织关系关联特征 组织技战术 (TTP)特点、常用漏洞、常用攻击手法、持久化方法、攻击习惯、命名规则习惯、关联IP地址、关联域名、关联URL	[9, 30, 81, 82, 91, 96, 112, 124]	能有效溯源攻击者身份、适用于产业界、计算开销相对较小、具有一定因果联系、原理简单	过度依赖专家经验、需要大量先验知识、鲁棒性较差、较难溯源未知的攻击事件、无法解决无文件攻击或离地攻击、智能化程度低
攻击模式级别特征	针对攻击模式溯源所提出的新型特征, 旨在通过对黑客组织攻击模式的上层抽象来匹配实现溯源追踪的特征集合	网络杀伤链模型、ATT&CK模型[43]	[29, 51, 86, 134]	能弱关联相似的APT组织或区域、鲁棒性扩展性较好、缩小语义鸿沟	更倾向于辅助溯源工作且无法准确定位组织、误报率高、系统开销大、特征描述困难
对抗逃逸级别特征	是指对抗逃逸、伪装、欺骗、混淆、环境模拟等操作所生成的特征集合	① 环境信息 硬件信息、编译器信息、虚拟机信息、软件信息、沙箱信息、控制信道 ② 对抗信息 混淆方式、加密算法、编码算法、免杀方式、正则表达式、漏洞信息 (如服务、端口等)	[111, 135, 136, 137]	能辅助对抗逃逸伪装操作、鲁棒性较好、具备一定抗混淆能力	环境和对抗信息较难捕获、与溯源任务存在异质性、粒度较粗、误报率高

由表 5 可知, 5 类级别的溯源特征能有效覆盖不同场景下的 APT 攻击, 通过多种特征形式以供溯源模型学习并最终实现对区域、组织、IP 地址、攻击者和攻击模式的溯源. 具体描述如下.

1) 代码基因级别特征旨在标识恶意代码的关键功能和行为, 动态特征可通过虚拟环境 (如 Cuckoo 沙箱、QEMU 模拟处理器) 和内存执行来动态提取, 静态特征可以利用静态分析工具 (如 IDA、010Editor)、计算文件的熵值或转换成灰度图来提取, 以样本内容为主, 最常用的特征是字符串、控制流图、API 序列和系统调用行为.

2) 系统日志级别特征主要是由系统审计日志中的事件和告警所生成的特征, 常见内核级审计工具包括 Windows 事件跟踪器 (event tracing for Windows, ETW) 和 Linux 系统的 Auditd 工具[133]. 随着溯源图被应用于 APT 攻击检测和溯源中, 系统日志级别特征扮演着越来越重要的角色, 通过关联因果依赖事件构建溯源图或知识图谱能有效追踪目标. 常见的特征包括进程、文件、Socket 网络通信、注册表和审计日志事件 (含读、写、执行、连接等操作).

3) 关联分析级别特征主要结合关联分析和社会工程学技术实现 APT 攻击溯源, 该级别特征在产业界和真实的安全防御系统中较为常见. 首先能通过社会工程学关联特征实现组织或攻击者的溯源, 譬如 Checkpoint 公司[124] 通过邮箱、Twitter 和 GitHub 账号关联溯源僵尸网络攻击的作者; 其次可以关联从 APT 样本中捕获的 IOC 特征, 不同于代码基因级别的 IOC 特征, 该类别是安全公司建立特有安全知识库, 通过 C&C 地址、Mutex 或载荷

HASH 值实现关联分析, 比如安天公司的追影威胁分析系统能有效追踪南亚次大陆的白象组织[81]; 最后通过技战术及关系能有效追踪 APT 组织的来源[30].

4) 攻击模式级别特征旨在弥补传统溯源方法无法有效识别 APT 组织采取模拟、伪装或组装恶意代码所发起的网络攻击来源的缺陷. 该特征会在底层攻击事件和顶层攻击组织间建立抽象的攻击模式 (即映射至网络杀伤链模型或 ATT&CK 模型[51,134]), 通过该方式缩小语义鸿沟, 溯源模型在该类特征的基础上学习并刻画不同组织的攻击模式, 最终利用不同组织攻击模式的差异来实施溯源.

5) 对抗逃逸级别特征是为防止攻击组织和黑客消除溯源线索所提出的特征. 该类特征主要对抗逃逸、伪装、欺骗、混淆、环境模拟等隐蔽性较高的操作, 以加壳混淆和系统环境信息检测为主, 譬如通过捕获环境信息来判断攻击载荷是否在虚拟机中生成, 或判断攻击是否存在伪装成其他组织的可能, 以及分析 APT 攻击是否由 0day 漏洞所发起[136].

在提取溯源特征后, 需要对溯源特征进行处理. 本文囊括了常见的 5 个关键步骤, 分别是数据预处理 (data preprocessing)、特征选择 (feature selection)、特征对齐 (feature alignment)、特征融合 (feature fusion) 和向量表征 (vector representation). 其中, 数据预处理是通过数据清洗和标准化处理得到质量更高的数据, 譬如吕杨琦等人[79]将内存地址统一标准化为 MEM, 指令示例为“mov esi, [esp+4]→mov esi, MEM”. 特征选择从大量溯源特征中选择更具有代表性的特征集合, 从而提升溯源模型的预测结果, 譬如张杰等人[136]结合逃避型恶意样本特点提取常见的 3 类逃避类型 API 函数特征, 包括调试器检测、沙箱检测和虚拟机检测特征. 特征对齐和特征融合旨在将具有歧义且表征目标一致的特征或关系进行对齐融合, 能有效提升匹配精度, 譬如 Milajerdi 等人[52]将审计日志构建的溯源图 (provenance graph) 和威胁情报关联构建的查询图 (query graph) 进行特征对齐和关系匹配, 最终实现对 APT 攻击的检测, 该工作设计了两种类型对齐, 分别是节点对齐 (node alignment) 和图对齐 (graph alignment). 向量表征旨在将特征表示或嵌入成溯源模型能学习的数学矩阵或向量形式, 常见方法包括 TF-IDF、Word2Vec、Asm2Vec、DeepWalk 等, 譬如 Steven 等人[77]提出结合语义和词汇关系以及二进制函数的向量表示, 利用二进制代码表示学习模型 Asm2Vec 实现二进制搜索和代码克隆检测.

3.4.2 溯源方法

经过溯源特征抽取和溯源特征处理后, 接下来将构建溯源模型, 它也将决定整个 APT 攻击智能溯源任务的最终结果. 本文结合学术界和产业界相关研究, 将溯源方法划分为 4 类, 分别是基于规则匹配和关联分析的溯源方法、基于机器学习的溯源方法、基于深度学习的溯源方法以及基于溯源图和知识图谱的溯源方法.

(1) 基于规则匹配和关联分析的溯源方法

传统 APT 攻击溯源主要借助专家经验实现, 通过关联溯源规则、安全知识库 (含 IOC、黑客人员信息) 和 APT 组织技战术特点来匹配溯源特征, 再利用关联分析、统计学相关算法或同源分析技术剖析 APT 攻击, 最终判定溯源对象身份和网络攻击来源.

该类方法属于早期的 APT 攻击溯源方法, 包括规则匹配、关联分析和同源分析, 主要应用于产业界的安全产品中. 规则匹配旨在通过制定规则或正则表达式来匹配特征, 进而追踪攻击来源, 安天安全研究与应急处理中心[81]利用逆向分析所提取恶意样本的 C&C 地址和编译器时间戳来溯源白象组织. 关联分析是利用社会工程学关联特征、样本 IOC 关联特征、组织关系关联特征构建溯源特征库, 并建立相似性度量方法或关联分析算法实现溯源, Kaspersky[82]、FireEye[30]和 Checkpoint[96]分别通过动态行为、组织 TTP 特征、代码特点和攻击行为模式实现 APT 攻击关联分析溯源. 同源分析是基于特征的相似度计算 (如余弦相似度算法) 或聚类算法 (如 DBSCAN 算法) 实现同源判定[14], 相似分数越高, 其同源程度越高, 溯源预测结果越准确, 典型方法如模糊哈希[79]、BinDiff[138]和 DeepBinDiff[139]. 其中, BinDiff 是二进制同源分析工具, 其会根据每个函数的流程图 (含代码块数目、代码块之间的关系、子函数调用数目) 生成一个签名, 再通过签名、函数和基本块的匹配来计算样本间的相似度和置信度.

根据上述研究发现, 基于规则匹配和关联分析的溯源方法能在一定程度上溯源目标区域、组织和攻击者, 可以帮助安全工程师快速找到攻击的相似性和差异. 然而, 该方法的自动化和智能化程度较低, 且过度依赖专家知识和人工标注, 无法有效识别具有混淆、对抗或新型网络攻击 (如无文件攻击) 的来源. 此外, 该方法的鲁棒性、扩

展性和准确性较差, 后续逐渐与机器学习或深度学习结合来提升模型的性能.

(2) 基于机器学习的溯源方法

随着机器学习被广泛应用于各行各业, 科研工作者将其用来溯源 APT 攻击. 该类方法通过构建机器学习模型对溯源特征进行训练学习, 再对测试集中未知的网络攻击来源进行预测. 常见算法包括支持向量机 (SVM)、决策树 (decision tree)、随机森林 (random forest)、逻辑回归 (logistic regression)、K 近邻 (K-nearest neighbor)、朴素贝叶斯 (naive Bayes)、K-means、DBSCAN 等算法.

1) 随机森林是一个包含多棵决策树的分类器, 它采用投票方式决定输出类别. Wang 等人[95]通过构建随机森林分类器学习所抽取二进制文件的动静态特征, 以实现 APT 组织溯源. Caliskan 等人[83]利用反汇编和反编译器提取二进制代码的指令、控制流图特征来训练随机森林分类器, 最终实现对源代码作者的识别. Laurenza 等人[94]结合 APT 组织的静态特征构建孤立森林分类器并判断样本所属组织.

2) 支持向量机是一种经典的监督学习算法, 其核心思想是将数据在高维空间中构建一个满足分类要求的超平面以完成分类任务. 黄克振等人[33]构建 SVM 分类器来学习攻击事件的线索数据和威胁情报特征. Alrabaee 等人[31]归纳二进制文件作者相关特征并利用 SVM 算法溯源恶意软件的作者. Kumar 等人[140]针对 APT 恶意样本的隐蔽性提取动态和静态特征以及恶意软件渗透初始阶段的重要信息, 再使用 SVM 算法训练模型以完成 APT 恶意软件溯源任务.

3) 聚类算法根据特征的相似性将数据集中具有相似特点的成员聚合在一起. FireEye 公司[30]提取 APT 组织的文档主题摘要, 利用 TF-IDF 技术和聚类算法溯源黑客组织. Rosenblum 等人[105]通过提取二进制代码的内部结构、CFG 图、代码风格和函数调用特征, 再构建 K-means 模型来溯源作者. 乔延臣等人[108]利用 K-means 聚类算法学习 WinAPI 调用习惯, 最终完成恶意代码作者的同源判定. Hong 等人[119]基于聚类算法设计 Ctracer 系统并在企业环境中成功识别 C&C 会话和攻击来源.

与前面基于规则匹配和关联分析的方法相比, 机器学习算法具备一定程度的智能性, 这种半自动化的溯源方法在样本数量较少情况下具有较好的准确率和较低的误报率. 此外, 该类方法具备一定的可解释性, 通过特征关联能有效发现溯源不同对象的关键特征, 且算法原理简单、易于编程实现. 然而, 由于机器学习算法属于浅层学习, 其较难挖掘特征间的深层次语义信息和上下文依赖关系, 且在特征工程阶段仍然需要安全专家进行标注和划分特征类型, 面对大规模日志、流量、样本和 APT 报告数据时, 其应用的扩展性和准确率会下降, 并且该类方法仍无法有效感知高隐蔽和高对抗攻击的来源. 因此, 如何有效从海量数据中准确识别出攻击来源, 第一时间感知不同类型的攻击, 减少对专家知识的依赖变得非常重要.

(3) 基于深度学习的溯源方法

下面详细描述基于深度学习的 APT 攻击溯源方法, 常见算法包括卷积神经网络 (CNN)、循环神经网络 (RNN)、自编码器 (AutoEncoder)、结合注意力机制的神经网络、图神经网络 (GNN) 和迁移学习 (transfer learning).

1) 卷积神经网络能有效提取局部特征, 卷积核滑动能提高特征的利用率, 其最早被应用于计算机视觉领域, 通常由卷积层、池化层和全连接层组成. 在恶意代码检测中, 将二进制样本转换成灰度图像再构建 CNN 模型实现分类预测是典型应用案例[141]. 同样, 在网络攻击溯源中也有应用. Alrabaee 等人[84]基于 CNN 模型构建二进制作者溯源工具 BinEye, 能够借助较少的专家知识学习作者编程风格追踪作者. 图 8 详细展示了面向恶意二进制代码的作者智能溯源方法, 二进制代码经过反汇编、反编译和代码解析生成指令特征、词汇特征、AST 特征和 CFG 特征, 最终构建 CNN-BiLSTM 模型实现作者去匿名化任务[83,104]. 文献 [142] 利用一种深度学习组合模型 (含 MLP、CNN 和 LSTM) 来检测网络攻击流量并有效识别 APT 攻击的 IP 地址.

2) 循环神经网络通常包括长短期记忆网络 (LSTM) 和门控循环单元 (gated recurrent unit, GRU), 其能更好地处理序列向量, 通过记忆单元捕获长距离依赖信息, 并解决梯度消失和爆炸问题. 李昂[32]通过提取攻击组织的代码基因特征并设计融合自注意力机制的双向循环神经网络完成 APT 组织溯源. Shang 等人[143]构建 CNN-LSTM 模型学习恶意软件和 C&C 服务器间网络流量的共享特征, 从而发现未知的 APT 攻击及 C&C 隐蔽通道. Wei 等

人[144]通过提取恶意样本的 API 系统调用特征, 利用 LSTM 和注意力算法来表示特征向量, 最终构建迁移学习模型实现 APT 组织溯源.

3) 图神经网络旨在利用神经网络来学习图结构数据, 并挖掘图结构数据中的特征和模式, 最早由 Marco 提出[145]. 由于 GNN 包含丰富的节点和关系信息, 能捕获更深层次的语义知识, 且具有强大的非线性拟合能力, 因此其在多个领域均取得良好性能. 李腾等人[97]提出一种基于图卷积神经网络的攻击组织溯源方法. Li 等人[28]提出一种基于注意力机制的图神经网络模型, 能在系统和网络层面捕获 APT 攻击的全生命周期行为并评估了模型的性能.

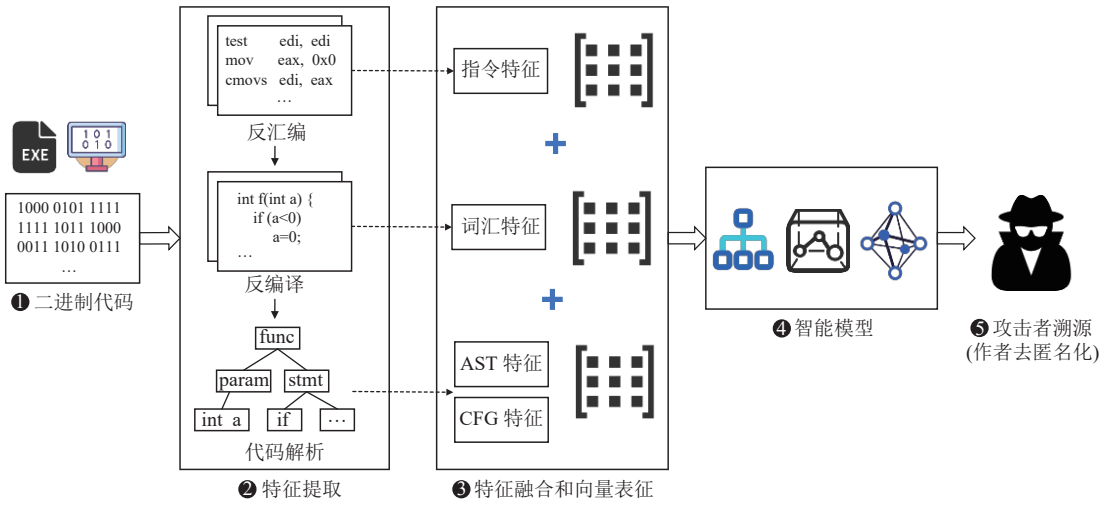

图 8 面向恶意二进制代码的作者智能溯源方法

根据上述研究发现, 基于深度学习的溯源方法能更好地捕获语义知识和长距离依赖关系, 能在少量标注的情况下取得良好的准确率和 F_1 值, 并且具有更好的鲁棒性和扩展性, 甚至有效挖掘未知或新型 APT 攻击的溯源特征从而追踪目标. 此外, 该类方法通常会与自然语言处理技术结合使用, 面对大规模数据也能取得较好效果. 然而, 该类方法的可解释性不强, 基于图神经网络和迁移学习的模型尚不成熟, 计算复杂性较高, 仍待完善.

(4) 基于溯源图和知识图谱的溯源方法

图 9 展示了面向审计日志的 APT 组织智能溯源方法整体流程, 通常包括数据预处理、特征提取、关系图谱构建、攻击社区发现 (或关系图谱优化)、APT 组织分类 5 个关键步骤. 李腾等人[97]抽取多源审计日志特征构建关系图谱, 利用图卷积神经网络和 Louvain 社区发现算法实现 APT 组织溯源. 黄克振等人[33]引入网络攻击事件溯源本体模型, 通过提取攻击事件线索和安全威胁情报特征来构建溯源关系图, 最后利用图嵌入和随机游走算法映射特征向量并追踪攻击组织来源. 此外, 将溯源图和知识图谱作用于其他类型的特征并开展溯源追踪的工作也逐渐产生. Han 等人[85]基于 APT 恶意样本的系统调用行为来构建溯源知识图谱, 最终实现 APT 组织溯源任务. Berady 等人[112]从攻防两个角度抽取 TTP 技术、组件和攻击痕迹特征, 再构建关系图谱实现对 APT29 组织攻击事件的识别. ProTracer[86]和 TRACE[123]是两个典型基于溯源图的跟踪系统, 通过提取内核代码和程序单元依赖关系构建溯源图, 最终实现对攻击者的调查取证和溯源追踪.

通过深入地归纳和对比分析发现, 相比于先前的 3 种方法, 基于溯源图和知识图谱的溯源方法具有更好的自动化和智能化水平. 该方法会将溯源特征 (如审计日志) 与攻击模式相结合来构建关系图谱, 以缩小语义鸿沟, 并且会按照因果依赖关系或时间顺序有效地关联攻击事件, 识别出隐蔽性较高的 APT 攻击事件, 甚至通过事件痕迹或子图匹配来溯源攻击对象, 还原攻击场景, 实现语义推断. 因此, 学术界和产业界目前均在溯源图和安全知识图谱领域开展深入研究, 以期望设计出鲁棒性更高、智能化更强且适用于大规模安全数据的模型或产品, 尽可能地对在野攻击开展实时监测, 第一时间感知未知的 APT 攻击并迅速定位出攻击来源, 给出攻击者画像.

综上所述, 我们详细介绍了 4 类智能溯源方法, 分别是基于规则匹配和关联分析的溯源方法、基于机器学习

的溯源方法、基于深度学习的溯源方法以及基于溯源图和知识图谱的溯源方法, 并全面梳理各类方法的原理、代表性工作及发展历程, 同时对比分析各类方法的优缺点. 此外, 溯源模型构建成功之后, 还需要和溯源评分与计算阶段相结合 (详见图 6 的第 5 个阶段), 通常包括相似度匹配、阈值计算、代码基因识别、同源分析、溯源评分及排序、溯源判定 6 个关键步骤, 它们能够有效实现对 5 种溯源对象的识别. 其中, 相似度匹配和同源分析主要针对关联分析级别特征, 如 FireEye[30]、Checkpoint[96]、Alrabaee 团队[31, 107]和 BinDiff[138]溯源工作. 代码基因识别主要抽取关键代码特征作为溯源的线索, 并追踪攻击者身份, 譬如安天安全研究与应急处理中心[81]、Checkpoint[124]的溯源工作. 阈值计算和溯源评分及排序主要解决审计日志数据爆炸问题, 通过对溯源图剪枝优化来提升模型的计算性能, 典型的工作是 HOLMES 系统[51]. 最后通过溯源判定给出模型预测的溯源对象. 自此, 本文已完成 APT 攻击智能溯源的综述研究, 并详细给出各阶段的任务描述及代表性工作.

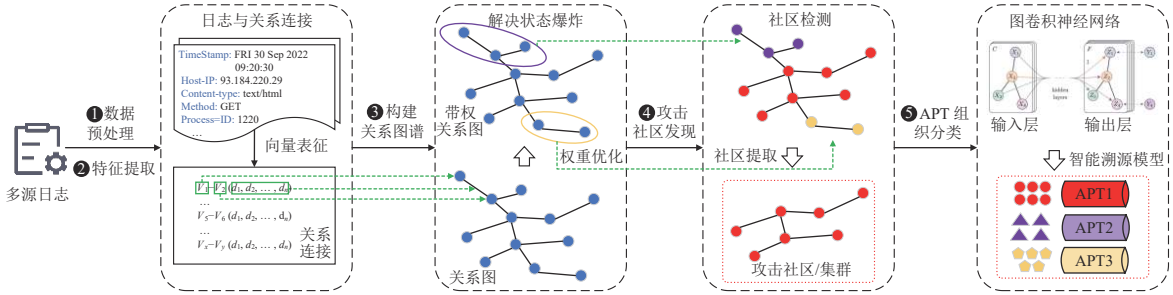

图 9 面向审计日志的 APT 组织智能溯源方法

3.5 溯源对抗

尽管学术界和工业界分别对网络攻击、定向网络攻击和恶意代码开展部分溯源研究, 上述描述也介绍攻击溯源困难的缘由. 然而, 随着网络攻防博弈和 APT 攻击技术的不断发展, 现有溯源技术仍然存在诸多挑战. 如图 10 所示, 攻击者通过隐藏和伪装手段将正常网络流量和系统日志信息进行处理, 尽可能地消除溯源痕迹和关联线索, 本文将其称为溯源对抗. 经过溯源对抗处理, 识别者 (溯源引擎) 将错误地定位和追踪目标对象.

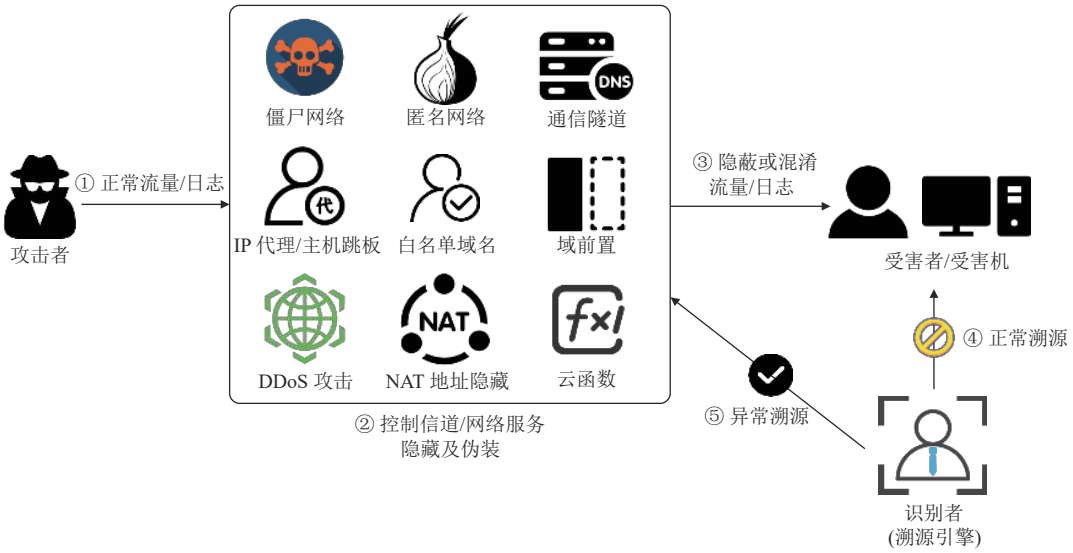

图 10 APT 攻击溯源对抗流程

当前最前沿且常用的溯源对抗方式包括 9 种, 包括僵尸网络、匿名网络、通信隧道、IP 代理/主机跳板、白名单域名、域前置、DDoS 攻击、NAT 地址隐藏和云函数. 一方面, APT 组织会利用匿名网络、通信隧道、主机

跳板技术隐藏攻击流量, 导致现有检测系统较难识别出真实的攻击地址; 另一方面, APT 组织通过僵尸网络发起 DDoS 攻击, 导致目标地址较难被准确定位, 或者通过白名单域名、域前置、NAT 地址隐藏和云函数技术来伪装或隐藏网络攻击, 从而欺骗溯源引擎. 此外, 由于恶意代码和攻击方式不断演化, 其通常会结合加壳、混淆、加密、反调试技术来对抗恶意代码检测, 从而导致传统方法的误报率和漏报率升高, 攻击者的真实身份和攻击发起地址较难溯源, 且以无文件攻击或 0day 漏洞的高隐蔽攻击给网络空间安全带来严重威胁[5,14,136]. 因此, 如何有效检测溯源对抗手段, 避免伪装、混淆、欺骗等技术的干扰变得至关重要, 这也是未来智能检测和溯源的研究重点.

最后, 本文结合溯源对抗知识, 从以下 5 个层面归纳 APT 攻击智能溯源存在的困难及挑战.

1) 在技术层面: APT 组织通常会利用对抗、伪装、欺骗等手段隐藏攻击发起的真实地址 (含物理区域和 IP 地址), 亦会采用混淆、加密、加壳、离地攻击等技术处理网络流量或恶意代码以躲避检测, 从而造成溯源引擎较难识别出攻击者真实身份, 无法有效将现实世界与虚拟网络的组织人员信息关联, 进一步增加网络攻击事件溯源的难度. 随着网络攻击事件增加, 溯源分析维度和数据量也不断扩充, 已有工作需要借助专家经验来制定规则, 在特征挖掘和关联分析上存在一定的局限性, 亟需自动化的方法来辅助 APT 攻击溯源研究. 此外, 现有技术以被动溯源模式为主, 为实现实时精准主动溯源的要求和需求, 结合前置防御和主动防御 (如蜜罐) 技术的主动溯源将成为未来研究的重点.

2) 在领域层面: 现有研究倾向于面向非定向网络攻击, 缺少主动获取攻击关键信息并实施智能化溯源的工作. 一方面产业界忽视深度学习和知识图谱等新兴技术对溯源特征深入挖掘的影响, 缺乏对已有攻击知识库的有效关联; 另一方面学术界缺少大规模 APT 攻击事件相关的数据, 忽视产业界业务和实效性需求. 此外, 学术界和产业界尚未有效结合 APT 攻击完整生命周期, 开展检测、溯源和推理相关联的研究. 尽管检测、溯源和推理之间的依存关系和界限尚未明确, 但将三者关联起来开展研究将进一步推动整个领域的发展, 这也是本文的追求和目的.

3) 在智能化层面: 传统溯源技术过度依赖专家知识且需要大量安全技术人员开展手工分析, 自动化和智能化程度较低[12], 且恶意代码家族和网络攻击形式演变较快, 新的变种和特征不断生成, 这会导致传统溯源技术较难识别新型未知的攻击并无法溯源其攻击来源. 同时, 攻击通常会潜伏在正常流量或常规操作中, 如何从海量信息中快速精准地提取攻击关键特征将变得至关重要, 如何通过智能化算法从多维复杂的信息中挖掘出溯源特征, 发现其背后的关联, 将是未来研究的重点, 当然 APT 攻击的溯源和推理研究也存在诸多挑战, 自动化和智能化溯源还处于起步阶段.

4) 在协议和架构层面: 早期的 TCP/IP 协议架构未考虑攻击源和隐私安全问题, 不能有效判断数据包的来源和威胁性, 这也加深了网络攻击溯源的困难. 此外, APT 组织通常具有极强的专业性, 攻击者身份和 IP 地址信息通常会被隐藏, 再加上现有内外网数据交互过程中缺乏系统性的防御措施, 这会导致多个突破口出现, 进一步使得溯源系统无法实现对 APT 攻击各个阶段的完整追踪.

5) 在社会和法律层面: APT 组织和黑客人员会利用社会工程学和关联分析技术挖掘被攻击对象的身份信息, 甚至建立攻击画像, 再结合心理学、工作及环境需求来伪装攻击, 比较典型的是鱼叉式钓鱼邮件和水坑攻击. 此外, APT 组织通常受到特定机构或部门的支持, 攻击组织所开发和使用的渗透工具和恶意代码也不尽相同, 模拟和伪装成其他组织来发起特定攻击事件也逐渐增多, 当前法律在互联网和隐私安全保护尚未考虑周全, 这也给 APT 攻击溯源带来挑战.

3.6 小　结

本节主要介绍面向 APT 攻击的智能溯源研究, 基于学术界和产业界关于 APT 攻击溯源的工作开展详细的综述总结和归纳分析. 首先, 本文给出 APT 攻击溯源的整体框架, 包括溯源数据采集、溯源特征抽取、溯源特征处理、溯源模型构建、溯源评分与计算、溯源对象确定 6 个关键步骤. 其次, 本节结合 5 类溯源对象 (区域、组织、攻击者、地址和攻击模式), 通过综述方式详细比较不同方法之间的优缺点以及未来研究的趋势. 再次, 给出 APT 攻击智能溯源的定义, 通过溯源机理的梳理进一步明确溯源任务的重要性. 接着, 本文详细归纳常见的 5 类溯源特征, 包括代码基因级别特征、系统日志级别特征、关联分析级别特征、攻击模式级别特征和对抗逃逸级别特征, 并将溯源方法划分为 4 类, 分别是基于规则匹配和关联分析的溯源方法、基于机器学习的溯源方法、基于深度学

习的溯源方法以及基于溯源图和知识图谱的溯源方法. 最后, 本节介绍溯源对抗知识以及 APT 攻击智能溯源存在的困难及挑战.

图 11 按照时间顺序和溯源技术发展路线列出了所综述的主要方法. 传统方法倾向于利用规则匹配和关联分析实现, 其发展时间较早且以工业界为主; 随后衍生出基于机器学习和深度学习的溯源方法, 该类方法追求与前沿的神经网络技术融合 (如图神经网络和迁移学习). 随着溯源图和知识图谱出现, 应用该技术实现 APT 攻击溯源的工作出现, 该类方法能更好地关联深层次语义知识, 并且与大规模审计日志场景联系更紧密.

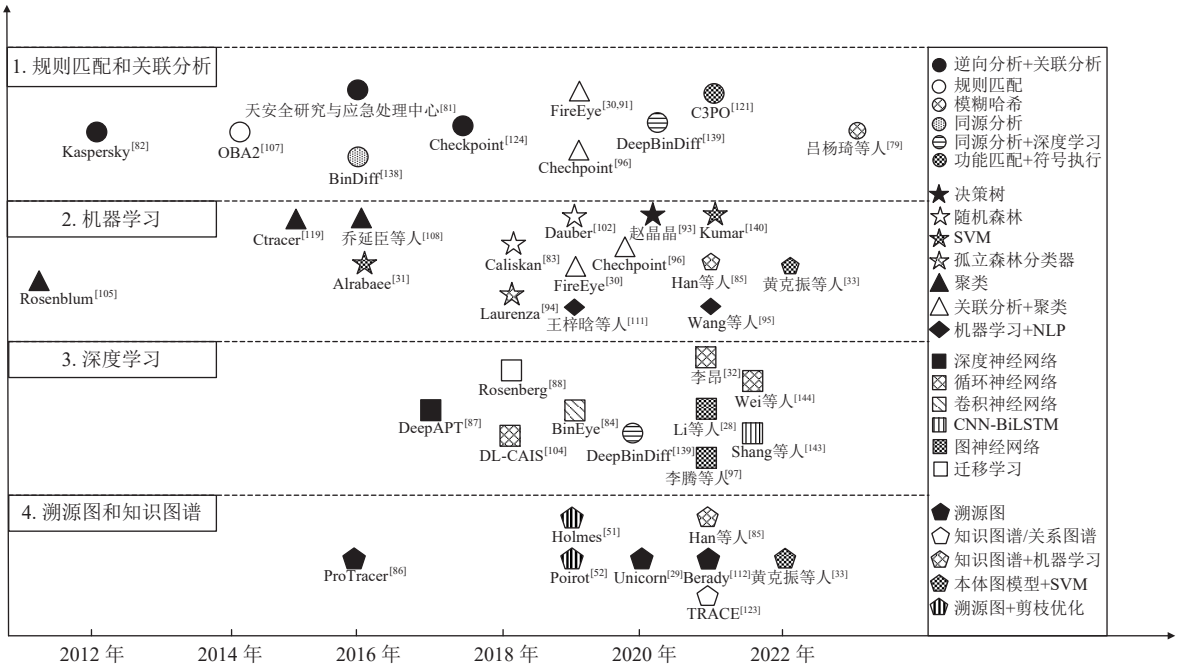

图 11 APT 攻击溯源主要方法及分布

总而言之, 整个学术界和产业界尚未形成覆盖 APT 攻击完整生命周期的检测、溯源和推理研究体系, 缺乏完整系统的智能化溯源方案, 针对无文件攻击和利用 0day 漏洞实施内网渗透的新型攻击仍较难解决. 未来, 针对隐蔽性更高、对抗性更强、复杂程度高、可持续时间长的 APT 攻击溯源研究将是整个领域的重点和难点, 与溯源图、知识图谱和攻击推理结合的研究将有所增强, 智能化和自动化的 APT 攻击溯源方法亟需突破.

4 APT 攻击推理

推理 (reasoning 或 inference) 是由已知的判断或前提推断出结论的过程. 随着人工智能技术不断发展, 探索未知世界的知识推理及智能问答等技术逐渐出现, 利用已知的知识去发现新的知识、预测潜在的事实或给出智能化的策略将成为新的研究热点.

在早期阶段, 王永庆[146]将推理定义为人类对各种事物的分析、综合和决策, 以及利用已知的知识寻找蕴含的事实或归纳出新的事实的过程. Tari[147]认为知识推理是基于特定的规则和约束, 从存在知识中获取新知识的过程. 随后, 官赛萍等人[21]从知识推理的基本概念出发, 详细综述面向知识图谱的知识推理方法的最新进展, 并将其划分为单步推理和多步推理, 方法主要包括基于规则的推理、基于分布式表示的推理、基于神经网络的推理以及混合推理. 此外, 知识推理逐渐被应用于各个领域, 包括智能问答[148]、生物医疗[149]等. 同样, 在网络安全和 APT 攻击领域逐渐开展推理研究, 通过智能方法推理攻击意图和行为[150,151].

基于此, 本文将开展面向 APT 攻击的智能推理研究, 详细综述和归纳现有方法的优缺点. 其中, 本文将 APT

攻击智能推理定义为利用机器学习、深度学习或知识图谱等智能技术去学习或关联已知的知识, 再构建模型来预测或发现潜在的 APT 攻击新知识的过程. 该任务能够有效预测攻击后续的意图或行为, 感知攻击路径、还原攻击场景并阻断后续的 APT 攻击. 通过全面地调研和总结, 本文将 APT 攻击推理划分为 4 个关键子任务, 分别是攻击意图推理、攻击路径感知、攻击场景还原、攻击阻断及反制. 整个框架如图 12 所示, 由数据采集、特征抽取、知识图谱或溯源图构建、推理计算和 APT 攻击推理 5 部分组成.

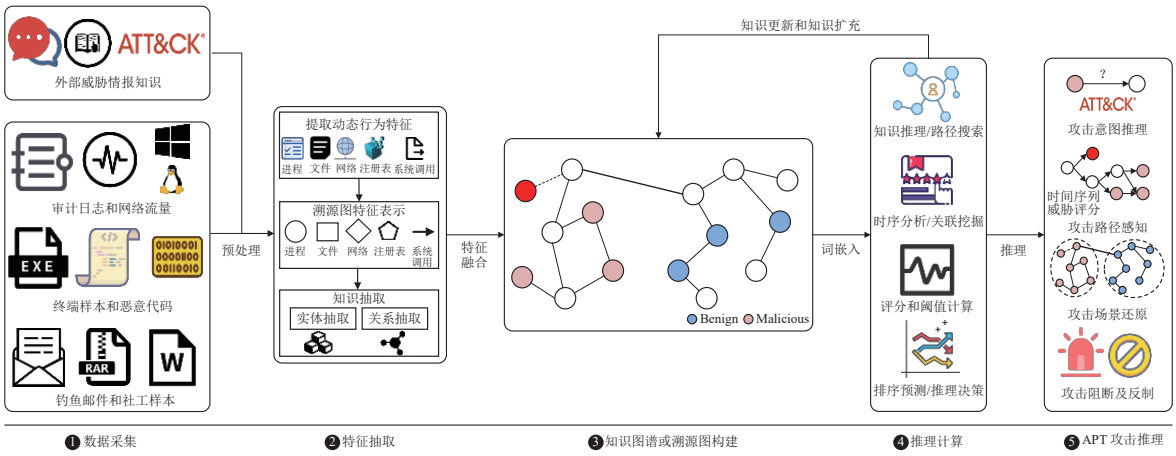

图 12 APT 攻击智能推理整体框架

数据采集包括流量侧和终端侧 3 种类型数据源, 分别是审计日志和网络流量、终端样本和恶意代码、钓鱼邮件和社工样本; 特征抽取包括行为特征抽取、实体抽取和关系抽取; 知识图谱或溯源图构建是结合抽取的知识构建初始知识图谱, 后续会不断更新和扩充知识; 推理计算旨在优化知识图谱或模型并实现最佳推理, 含知识推理和路径搜索、时序分析和关联挖掘、评分和阈值计算、排序预测和推理决策 4 个关键步骤, 结合时序关联 APT 攻击的行为路径; APT 攻击推理主要包括 4 个关键子任务. 后续章节将详细概述和总结各个子任务.

需要注意, 随着 APT 攻击的隐蔽性、对抗性、破坏性和逃逸性不断增强, 未知的攻击变体不断增多, 推理和理解 APT 攻击的意图和行为, 还原攻击场景和路径, 以及阻断和反制 APT 攻击均存在诸多挑战. 此外, 学术界和产业界针对 APT 攻击的推理研究仍处于起步阶段, 相关工作相对较少. 因此, 本节主要结合现有工作和作者多年从事相关研究的经验, 并对未来研究趋势的预测进行概述, APT 攻击推理作为 APT 攻击防御链的最后阶段至关重要.

4.1 攻击意图推理

攻击意图推理旨在构建知识图谱或智能模型来推断 APT 攻击的意图或后续攻击行为, 以及通过智能问答的形式实现网络攻击知识的搜索或给出对应的攻击策略. 在早期阶段, 研究者通过定义规范或规则匹配来推理安全知识, 为基础设施提供更好的安全保障[152]. Paja 等人[153]提出一种用于安全需求建模和推理的系统, 通过 STS-ml 建模语言来自动推理安全需求之间可能存在的冲突以及安全需求与参与者业务策略之间的冲突, 整个系统涵盖社会组织、信息资产、系统权限 3 个视图, 通过定义概念、关系和约束来推理安全要求 (是否规范), 最终在电子政务案例中评估该系统的性能. Peralta 等人[154]利用 SOLj 安全操作语言构建传感器网络的逻辑框架, 利用逻辑演算和推理规则来评估系统资源的安全性和完整性. 文献 [155] 提出一种基于高性能计算的移动恶意软件推理算法, 通过高性能模拟工具 EpiCure 和人口流动网络推理智能手机中恶意软件的传播趋势. 针对网络攻击信息被黑客组织广泛共享, Marin 等人[156]引入与黑客攻击活动相关联的注释概率时间规则和演绎学习 (谓词逻辑) 来预测即将发生的网络攻击事件, 该研究表明时间线在攻击推理中扮演着重要角色. 邢倩倩[157]提出一种基于智能规划的网络安全场景模型, 将网络场景的风险过程分析转换为安全规划问题, 利用 PDDL 语言推理各情景下的安全风险并生成攻击图. Karafili 等人[158]提出一种基于论证和演绎推理的推理机 (ABR), 通过分析 APT 攻击的技战术和证据信息来完成自动推理并实现取证分析, 譬如回答谁是攻击者、谁有实施攻击的动机问题. 公式 (2) 是推理规则示例,

表示实体 X 具有执行攻击的 Att 经济动机. 其中, $industry(T)$ 是一个背景事实, 说明 T 是一个行业 (工业); $target(T, Att)$ 是一个证据, 说明 T 的目标是 Att; 谓词 $hasEconMot(X, T)$ 是一个证据, 说明实体 X 从攻击行业 T 中获得经济利益; $contextOfAtt(econ, Att)$ 表示 Att 经济充裕; $specificTarget(Att)$ 指攻击特定目标是 Att.

$$hasMotive(X, Att) \leftarrow \begin{cases} target(T, Att) \\ industry(T) \\ hasEconMot(X, T) \\ contextOfAtt(econ, Att) \\ specificTarget(Att) \end{cases} \tag{2}$$

Ghosh 等人[159]利用恶意软件感染树 (malware infection tree, MiT) 和马尔科夫链模型来预测恶意软件行为, 以揭示恶意软件感染的过程, 譬如推断 Backdoor. Win32.Poison 恶意软件的文件和进程创建的可能性. 其中, MiT 的节点表示文件或进程, 边表示捕获的行为关系. 随后, 基于演绎学习、形式化描述或概率决策的攻击意图推理被应用到各类安全场景, 如智能家居设备意图推理[160]、云服务部署安全评估[161]、内存安全漏洞风险识别[162].

然而, 早期推理需要技术人员开展大量的手工标注或定义规则 (约束), 该类方法的自动化程度较低, 且缺乏深层次的语义知识关联, 无法实现对规模较大与隐蔽性较高的 APT 攻击推理, 也缺乏系统的攻击场景验证. 随着知识图谱、溯源图和本体技术的不断演化, 基于知识图谱和依赖关系的攻击意图推理工作逐渐增加. 针对安全知识图谱构建, Ma 等人[163]提出一种基于条件随机场 (conditional random field, CRF) 和 BiLSTM 的安全命名实体识别 (named entity recognition, NER) 模型, 能从非结构化文本中提取网络安全相关的概念和实体. 杨秀璋等人[164]提出一种融合实体识别和实体对齐的 APT 攻击知识自动抽取方法, 该工作结合 APT 攻击特点和 ATT&CK 框架定义 12 种实体, 包括 APT 组织、攻击装备、攻击手法、攻击漏洞、攻击事件、恶意软件家族和区域位置等, 再构建 Bert 和 BiLSTM-CRF 模型识别 APT 攻击实体, 最终生成不同 APT 组织的结构化知识. 经过实体识别和知识提取后, 能有效构建安全知识图谱, 这将为后续安全知识推理提供支撑.

在安全知识推理方面, Solic 等人[165]针对现有安全解决方案缺乏知识库的支撑, 提出一个评估信息系统安全解决方案的模型, 利用 OWL 本体构建知识库, 使用推理算法 (Dumpster-Shafer 理论) 和数学计算来评估各种安全问题的可能性和等级, 包括网络安全、软件和硬件问题、人为因素、安全策略和灾难恢复. Wu 等人[166]提出一种基于本体和图的安全评估方法, 通过本体将资产、漏洞和攻击等安全知识形式化表示, 再利用本体模型的推理能力来生成攻击图、量化攻击风险以及制定对应的缓解计划, 图 13 展示该研究所生成的攻击本体示例, 以攻击、设备、漏洞和组件实体为主. TINKER[167]是一种聚合网络威胁情报的框架, 通过半监督学习模型提取大规模威胁信息来构建安全知识图谱, 且能有效推理未知的威胁信息, 包括漏洞、恶意软件和作者之间的关联. Zhang 等人[168]设计一种基于推导规则的发现算法, 通过检测时间、语义和流程相关的属性来发现网络攻击事件之间的时间和因果触发关系, 再引入基于用户意图的安全策略和触发关系图精确定位高隐蔽的恶意软件活动. He 等人[169]提出一种新型攻击图模型以理解与推理 Spectre 和 Meltdown 高隐蔽攻击的行为, 它们能破坏用户应用程序所提供的内存隔离, 从而获得隐蔽通道泄露的推断执行信息, 该模型通过将攻击操作建模为有序依赖关系图来识别攻击变体, 并生成对应的防御策略系统解释防御决策的缘由. Alserhani[170]构建基于知识的因果警报模型来表示安全信息并推理多阶段攻击的行为. Lansley 等人[171]利用案例推理和深度学习技术来检测社会工程学攻击. Zhao 等人[172]提出一种基于异构信息网络的网络威胁情报框架 HINTI, 旨在建模异构 IOCs 之间的依赖关系和描述威胁事件概况, 再利用基于图卷积网络的威胁情报计算框架来发现知识, 将具有相似攻击偏好的攻击者聚类, 最终为不断变化的攻击威胁提供新线索. 张书钦等人[151,173]针对物联网与供应链会遭受 APT 攻击, 传统方法未有效将攻击者的战术、技术和攻击模式与安全知识关联, 提出物联网供应链风险分析本体 RIoTSCO, 通过设计物联网安全的推理规则, 利用本体表达物联网安全领域知识的语义关系, 最终生成自动化推理缓解措施以防御威胁事件, 并且该本体能描绘威胁事件上下游供应链全貌. 为对抗高级网络攻击, Zengy 等人[174]提出 Shadewatcher 系统, 结合网络安全的威胁检测与信息检索中推荐之间的结构相似性, 将“系统-实体”交互映射为“用户-项目”交互的推荐概念来识别网络威胁. 该系统由知识图谱构建、推荐模型生成、威胁检测和模型调整 4 部分组成, 通过图神经网络感知审计记录的深层次信息, 并基于审计日志构建溯源图来实现对不同攻击场景的网络威胁的自动化和自适应分析, 最后在两种 APT 后门攻击中开展警报预测评估.

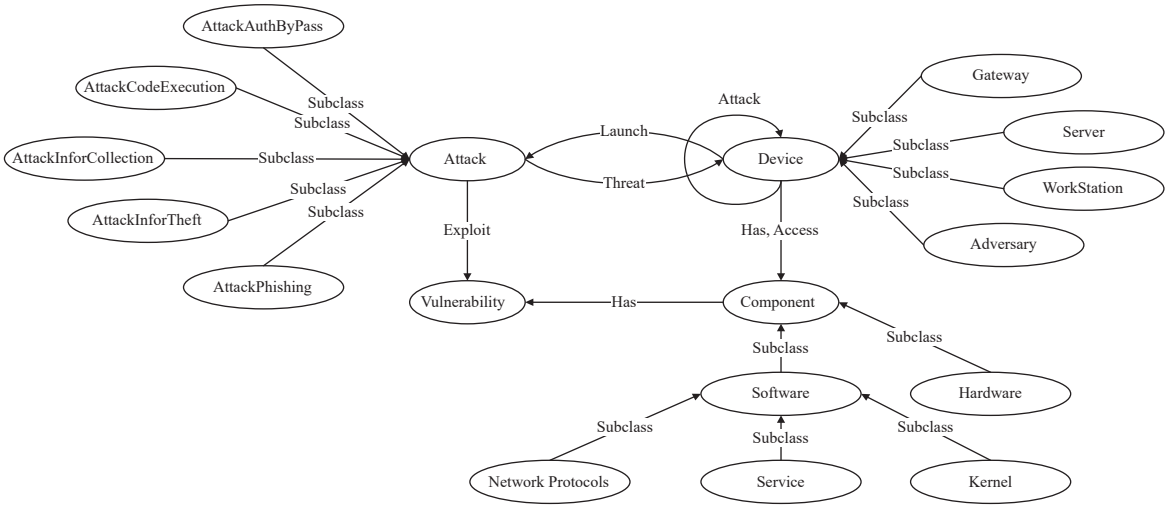

图 13 APT 攻击本体及依赖关系图谱示例

在攻击策略生成方面, QUASAR[175]是系统分析攻击和防御的框架, 能有效推理防御配置的覆盖范围, 生成新的攻击策略. Marin 等人[150]通过从暗网黑客论坛中收集攻击活动相关的语料, 再将其与现实世界网络攻击事件相关联, 挖掘 APT 攻击的逻辑规则来预测未来的网络攻击, 生成黑客组织及其计划的攻击策略. Zeng 等人[60]针对 APT 攻击分析面临着低级审计日志和高级系统行为的语义鸿沟问题, 提出 WATSON 系统, 其能够通过语义推理和聚合审计日志事件的上下文语义实现高级系统行为的自动抽取, 接着构建融合时间戳的安全知识图谱来表征攻击行为并生成不同行为实例 (子图) 的向量表示. Alsaheel 等人[176]针对 APT 检测系统存在误报和警报疲劳现象, 提出一种端到端攻击故事生成的框架 ATLAS, 利用自然语言处理、深度学习 (LSTM) 和因果关系分析构建序列模型, 并从因果溯源图中建立攻击和非攻击的关键模式, 最终生成抽象的攻击策略以及构建 APT 攻击故事.

4.2 攻击路径感知

攻击路径感知可以从攻击者和防御者视角进行描述, 前者是通过构建智能模型并结合攻击环境、目标信息开展攻击路径规划研究, 从而指导 APT 组织或攻击者实施精准攻击; 后者是详细分析攻击事件的依赖关系, 通过提取关键特征或依赖关系来预测潜在的攻击路径, 从而为后续攻击场景还原和攻击阻断提供支撑. 由于本文以 APT 攻击智能防御为主, 因此该部分主要从防御者视角归纳和总结攻击路径感知现有工作.

基于防御者视角, 攻击路径还原和预测能帮助安全人员更早地发现 APT 攻击, 协助防御者理解攻击的战略意图和后续行为, 从战术层面优化防御体系. 在早期阶段, Sigholm 等人[177]提出一种网络反情报的框架, 通过检测 APT 攻击活动泄露的信息来生成指纹, 再利用反情报传感器跟踪泄露数据的位置和传输路径. Zimba 等人[36]提出一种基于贝叶斯网络的云计算攻击路径识别方法, 旨在从 APT 攻击多个来源中寻找到最短的攻击路径, 并利用 WannaCry 勒索软件攻击来评估方法的有效性. Jia 等人[178]借助攻击树 (attack tree) 检测 APT 攻击, 通过攻击行为相关性分析来捕获高级可持续威胁的攻击路径, 并有效识别出软件定义网络中是否存在 APT 攻击. Akbar 等人[179]将 APT 攻击轨迹转换为图结构并构建深度学习模型 (GraphSAGE 和 OAML) 识别攻击路径和策略. 此后, 陈伟翔[180]针对当前研究侧重于寻找攻击特征检测 APT 攻击, 忽视预测攻击路径的问题, 提出一种 APT 攻击路径还原及预测的方法, 通过提取恶意软件的关键基因序列, 再利用隐马尔可夫模型还原 APT 恶意行为链 (通过基因检测获得) 的攻击路径并预测后续攻击路径状态, 接着在 APT29、Lazarus、Turla 等家族样本中进行路径推理实验. 张耀方等人[181]提出一种大规模工控网络关键路径分析方法, 设计一种仅更新关键节点攻击概率的贝叶斯攻击图动态更新策略, 高效计算全图的攻击概率并分析关键路径. Li 等人[182]提出一种基于 Web 日志的网络攻击痕迹关联分析方法, 利用攻击事件包含的类型、时间、源地址、目标地址等关键信息构建描述模型, 为取证分析和攻击溯源提

供帮助. Stellios 等人[183]利用攻击树和风险评估来识别物联网关键网络的物理攻击路径.

然而, 上述方法主要运用概率统计、机器学习或攻击树技术预测攻击路径, 缺乏深层次的特征和依赖关系挖掘, 忽视了攻击图、知识图谱和攻击事件因果关系对未知攻击路径的感知, 较难准确识别攻击技术不断演化的 APT 攻击的关键路径. 基于此, 融合知识和攻击图的方法逐渐被提出. 由于攻击图 (attack graph)[184]能够通过攻击状态转移对多步攻击行为关联进行建模, 因此被用于预测网络攻击路径. 胡浩等人[185]提出基于吸收 Markov 链的网络入侵路径预测方法, 将完整攻击图映射为吸收 Markov 链, 给出基于通用漏洞评分标准的状态转移概率度量方法, 结合攻击状态节点的访问次数和路径长度的期望值来预测入侵路径. 余洋等人[186]采用混合路径攻击图和多目标优化理论来计算 0day 漏洞利用风险, 并生成路径覆盖率更好的防御方案. Sadlek 等人[187]将杀伤链模型与攻击图结合, 通过对攻击者的行为序列进行建模来识别 APT 攻击路径和威胁场景. 翟海霞等人[188]设计一种基于 T_NAG 模型的路径预测方法, 通过改进攻击图和实时行为轨迹来区分攻击者能力并预测攻击路径.

此后, 基于依赖关系和知识图谱的方法出现. Xiong 等人[35]提出一种基于状态跟踪和检测的 CONAN 框架, 其能快速收集日志事件、调用堆栈和 API 序列, 提取数据之间的语义知识, 再将每个进程和文件都表示为类似有限状态自动机 (FSA) 结构, 最终利用智能策略来重构 APT 攻击链与高效检测 APT 攻击的可疑行为, 该研究在精度和效率方面明显优于先前的工作. 孙澄等人[189]提出一种基于知识图谱的 0day 攻击路径预测方法, 通过抽取网络安全领域概念及本体来构建网络防御知识图谱, 利用基于路径排序算法的知识图谱推理方法挖掘目标系统中可能发生的 0day 攻击并生成攻击图. Dai 等人[190]构建 Patrol 系统来推理 0day 漏洞的攻击路径, 通过捕获操作系统之间的依赖关系 (含调用) 来构建系统对象依赖关系图, 再利用入侵特征前向和后向跟踪来识别可疑传播路径. 卢嘉中[191]针对 APT 攻击路径较难还原的问题, 提出一种基于流量和网络设备日志的多阶段 APT 网络入侵检测算法, 通过改进受限波尔兹曼机 (RBM) 算法来增加检测面并还原 APT 网络入侵路径, 从而反映网络设备的状态及异常行为, 更好地实现 APT 攻击路径回溯和攻击重建. Zebra[192]是基于攻击模式搜索和因果依赖关系追踪的系统, 通过审计日志高效调查攻击并揭示攻击序列. Kim 等人[193]研究具有规避功能的网络钓鱼 URL 的特点, 提出一种基于网络的推理方法来检测逃逸型网络钓鱼 URL, 利用 URL、域名、IP、服务器和字符串构建异构网络, 再通过定制信念传播算法和边缘分配机制来实现相邻节点和关系的推理, 最终推断未知的网络钓鱼 URL.

需要注意, 其他安全防御领域也存在攻击路径感知和识别的工作, 譬如漏洞攻击路径挖掘[194,195]、利用动态路径标识阻止 DDoS 攻击[196]、利用移动轨迹跟踪保护隐私[197]等. 此外, 部分工作开始研究如何解决溯源图状态爆炸问题和优化攻击路径, 从而更好地感知网络攻击的关键路径[198].

4.3 攻击场景还原

APT 攻击场景还原旨在提取攻击的关键特征或依赖关系, 利用已识别的攻击路径和战术意图来重构 APT 攻击的场景, 感知攻击技战术和过程, 从而更早地发现 APT 攻击并实施预警, 为后续攻击阻断和反制提供帮助. 然而, 由于 APT 组织会利用对抗、逃逸、隐蔽等手段不断改进恶意软件, 且需要通过正在发生或未发生的攻击行为来还原攻击场景, 因此该工作的难度极大且存在诸多挑战, 相关研究工作也较少.

潘亚峰等人[40]详细介绍攻击场景重构的基本概念, 总结基于经验知识、因果关系、语义相似性和机器学习的 4 类方法, 旨在更好地从大量告警日志数据中, 依据特定攻击行为模式和语义知识, 分析数据间的关联关系还原攻击过程. 马琳茹等人[199]提出一种基于属性层次树的相似度隶属函数定义方法, 利用模糊聚类算法实现入侵检测告警关联图的重构. 伏晓等人[200]提出一种层次化入侵场景重构方法, 通过报警关联技术生成抽象的攻击步骤, 再利用攻击特征和依赖追踪技术重构各步骤的行为细节, 最终利用重构映射获得完整的入侵行为图. Tajima 等人[201]将攻击跟踪器应用于 APT 攻击, 通过对网络设备建模来模拟信息系统的行为, 利用 gRPC 接口模拟 APT 攻击的特定场景. Dain 等人[202]将多个入侵检测系统生成的警报组合到场景中, 从而识别不同类型攻击场景, 该方法主要借助专家经验和攻击模板重构攻击场景. 王硕等人[203]提出基于因果知识网络的攻击场景构建方法, 利用专家知识定义因果关系并结合数据挖掘刻画告警因果知识, 最终通过告警映射和最大后验估计原理重构攻击场景.

然而, 上述方法主要依赖专家经验知识、模拟跟踪器或因果关系来设计, 缺乏深层次语义关系的理解, 其对未知攻击和 APT 攻击场景的还原能力较弱[40]. 随着溯源图和知识图谱技术的蓬勃发展, 它们也被用于 APT 攻击场景还原. Hossain 等人[34]提出一种基于企业主机审计日志实时重构攻击场景的系统 SLEUTH, 首次利用溯源图和因果依赖关系来生成场景图 (scenario graph) 并重构 APT 攻击场景. 随后, Hossain 等人[204]针对依赖爆炸问题提出 Morse 系统, 通过引入两种标签衰减概念来移除良性依赖和过滤不相关路径, 从而构建更紧凑的场景图以识别高隐蔽的 APT 攻击. Hassan 等人[205]针对威胁警报误报问题, 提出一种利用威胁情报上下文和历史信息实现入侵检测的 NoDoze 系统, 其核心思想是通过警报事件的因果依赖关系图和异常分数来检测威胁, 并生成攻击场景的依赖关系. 随后, 该团队针对端点检测和响应 (EDR) 工具会产生大量误报、消耗资源等问题, 将战术溯源图引入该领域, 利用 EDR 生成的威胁警报之间的因果关系来构建 RapSheet 系统[134], 最终实现威胁检测并还原 APT29 组织的特定攻击场景. Fang 等人[206]提出 Depimpact 框架, 通过分析 POI 事件高度相关的依赖关系来提取关键边和代表性的攻击序列, 再结合反向传播和依赖影响过滤未发现的边, 以识别依赖关系图的关键组件. Milajerdi 等人[52]通过关联外部公开的网络威胁情报和内部攻击的审计日志溯源图, 利用图匹配和图对齐算法实现对典型 APT 组织的攻击检测和场景还原 (如 OceanLotus).

此外, 针对先前方法未有效关联 APT 攻击的攻击链或生命周期, Milajerdi 等人[51]将 APT 攻击活动信息映射到杀伤链中, 构建高级场景图实现低层次日志事件到高层次场景行为的映射, 最终实现对 APT 攻击的实时检测和攻击场景还原, HOLMES 是经典的 APT 攻击检测及场景还原系统. 琚安康[207]针对现有研究缺乏对告警事件与攻击上下文关联关系动态刻画的问题, 通过扩展网络威胁过程模型, 提出基于级联攻击链模型的定向网络攻击场景重构方法, 将攻击事件显示映射到攻击链的不同阶段. Zhu 等人[208]通过提取 IOC 信息并映射到攻击链的 4 个关键阶段, 有效将手动攻击分析与大规模攻击测量联系起来, 实现恶意活动语义学习. 需要注意, 由于外部威胁情报或审计日志具有文本数据特性, 其可以通过自然语言处理的某些方法优化. Satvat 等人[209]利用自然语言处理从网络威胁情报报告中精确提取攻击行为, 再通过语义角色标注进行语义分析, 理解攻击行为关系并将非结构文本转换为溯源图. Shen 等人[210]提出 ATTACK2VEC, 旨在利用词嵌入来理解网络攻击的演变和恶意行为在攻击步骤中发挥的作用. 同时, 其他安全领域也存在攻击场景还原工作, 譬如移动端 Android 恶意软件行为重建[211]、C&C 网络重构[212]、利用在线学习的更新泄露信息推理和重建攻击场景[213]、利用多重信息融合来还原异常流量的攻击场景及提取攻击链[214]. 未来, 该任务仍是 APT 攻击推理的难点和重点.

4.4 攻击阻断和反制

攻击阻断和反制旨在利用已推理的攻击意图、感知的攻击路径和重构的攻击场景来阻断 APT 攻击, 甚至反向获取攻击来源信息并反制攻击者. 传统方法倾向于利用蜜罐、诱饵、欺骗技术实现入侵来源的追踪和反制. 然而, 这些方法主要针对广义的网络攻击或威胁, 无法智能化感知和反制高隐蔽的 APT 攻击, 并且相关研究较少且存在诸多难点. 本节将总结和分析现有攻击阻断和反制的两类典型方法, 具体内容如下.

● 基于欺骗式防御的攻击阻断和反制技术. Gartner 公司提出攻击欺骗 (deception) 的概念[215], 旨在利用欺骗手段来阻断网络攻击事件, 拖延攻击者的入侵并有效检测攻击行为. 通常而言, 防御者会构造欺骗环境或行为来误导后续攻击活动以实现安全防护. 贾召鹏等人[216]形式化定义网络欺骗, 依据欺骗环境构建方式将网络欺骗划分为掩盖、混淆、伪造和模仿 4 种. Cohen[217]提出第 1 款利用欺骗技术实施安全防护的工具 DTK (deception toolkit). Araujo 等人[218]通过构造诱饵来收集攻击信息和欺骗性文件, 最终利用重制的安全补丁来阻断网络攻击. 杨俊岭等人[219]分析美军智能欺骗和反制相关技术手段, 形成“数字隐身”和“数字伪装”概念. 为扭转攻防双方不对称局面, 蜜罐 (honeypot) 技术被提出, 其本质也是利用欺骗技术布置诱饵信息, 诱使攻击方实施攻击, 防御方依此捕获攻击信息、推理攻击意图和实施攻击反制[220]. 诸葛建伟等人[220]详细介绍蜜罐技术的起源及发展历程. 同时, 分布式蜜罐、蜜网 (honeynet)[221]、蜜场 (honeyfarm)[222]技术被提出, 旨在提高欺骗技术的交互和捕获能力, 降低被识别的风险. Zhan 等人[223]提出描述和分析蜜罐捕获的网络攻击统计框架, 能有效预测网络攻击行为. Akiyama 等人[224]构

建蜜罐来分析恶意 URL 重定向攻击并提出阻断对策. Irvene 等人[225]针对机器人远程访问攻击提出 HoneyBot 来溯源攻击者以及阻止攻击. 此后, 针对高隐蔽和高威胁的网络攻击欺骗式防御方法出现. Fan 等人[226]提出一种名为 HoneyDOC 的蜜罐架构, 通过诱饵、捕获和协调器 3 个模块捕获高质量的攻击数据. Horák 等人[227]构建融合博弈论模型的蜜罐来识别和阻断网络攻击的横向移动. Tian 等人[228]提出针对工业物联网 APT 攻击的蜜罐检测策略. Jafarian 等人[229]利用 IP 地址随机化方法将终端主机转变为不可追踪的移动目标, 以规避扫描、蠕虫传播和定向网络攻击. 然而, 该类方法维护与部署代价较高, 主动防御的灵活性较差, 且缺乏深层次的语义分析, 无法有效阻断隐蔽性较高、恶意代码变体较多以及利用 0day 漏洞的 APT 攻击.

● 融合知识推理和 APT 攻击生命周期的攻击阻断和反制技术. 该类方法将融合机器学习、深度学习和知识推理技术, 结合 APT 攻击生命周期和外部威胁情报来构建更加逼真和智能的欺骗或反制系统, 以实现攻击阻断和反制. 陈瑞东等人[5]归纳了当前 APT 攻击检测与反制存在的 4 类问题, 包括渗透防护脆弱、检测精度低、攻击范围取证困难和未知新型攻击响应慢, 提出 APT 攻击阶段性防御方案和溯源反制技术路线. 贾召鹏等人[216]提出网络欺骗的层次化模型, 结合网络杀伤链概念分析攻击各阶段采用的欺骗技术以保护系统安全. Dai 等人[230]提出基于 Snort 和 OpenFlow 的 APT 攻击启发式感应蜜罐平台, 通过提取攻击流量特征并结合 ATT&CK 框架阻断 APT 攻击. 此外, Veluchamy 等人[231]构建基于深度强化学习的蜜罐环境来实时检测和阻断 DoS 攻击. 陈晋音等人[232]提出面向基于强化学习的智能渗透攻击的欺骗防御方法, 实现对渗透攻击前中后阶段的欺骗防御. Yang 等人[233]利用差分博弈方法生成 APT 攻击的修复策略. Foureye[53]基于超博弈理论和网络杀伤链推理 APT 攻击各阶段的最佳策略, 利用防御欺骗技术来检测和阻断 APT 攻击. 宋宇波等人[234]提出一种基于拓扑分析的网络流量分流和阻断方法, 利用多种发现策略获取网络拓扑, 基于主机行为特征溯源网络攻击, 最终构建基于流表的报文实时过滤方法来阻断攻击. HoneyPLC[235]是一款高交互、可扩展且支持可编程逻辑控制的新型蜜罐, 可有效收集和识别针对工业控制系统的侦察工具, 以识别攻击者的策略并利用欺骗技术阻断攻击. RAD[236]是一种基于行为分析的 DDoS 攻击对抗策略统计机制, 能有效识别潜在攻击状态并缓解 DDoS 攻击. Qu 等人[237]提出一种轻量级的对策方案来缓解缓存污染、缓存投毒和兴趣报文洪泛 3 种攻击. Ahmad 等人[238]详细分析 APT 攻击各阶段的战术、技术和攻击执行动机, 接着提出一种基于态势感知的虚假信息模型来反制 APT 行动及运营者的决策. Singh 等人[239]详细分析 APT 攻击的步骤、场景和对策, 讨论未来围绕 APT 生命周期的防御框架.

总之, 上述两种方法能在一定程度上实现 APT 攻击阻断和反制. 然而, 随着攻防博弈增强, 黑客组织会利用欺骗、伪装、隐蔽、0day 漏洞等手段开展定向网络攻击, 先前方法的效果会严重降低. 基于此, 本文预测未来会出现更多利用知识推理和融合生命周期的反制技术, 其将结合先前的攻击意图、路径和场景开展精准反制. 表 6 详细展现 APT 攻击智能推理代表性工作的方法和特点, 具体内容包括推理方法、推理特征、攻击类型、贡献及特点、先验知识、真实场景、鲁棒性、实时检测等.

表 6 APT 攻击智能推理工作总结表

分类	相关工作	推理方法	推理特征	应用场景	任务及特点	关系图谱	未知推理	先验知识	真实场景	DNN	鲁棒性	实时检测
攻击意图推理	STS-ml[153]	推理建模语言 定义约束和规范	社会组织/信息资产/系统权限相关节点和关系	电子政务	设计STS-ml建模语义 检测安全冲突	√	×	●	×	×	L	×
	HINTI[172]	多粒度注意力机制 异构信息网络 GCN+知识图谱	网络威胁情报IOCs信息 (含设备/漏洞/平台等)	网络威胁	自动识别IOCs 攻击偏好聚集 网络威胁发现	√	√	●	√	√	L	×
	张书钦等人[151,173]	物联网供应链本体 多源知识推理/规则 ATT&CK矩阵	物联网供应链安全知识 (含漏洞/攻击技战术等)	物联网供应链	生成自动化推理缓解措施并描绘攻击全貌	√	√	●	√	×	M	×
	Shade-watcher[174]	知识图谱+溯源图 GNN+推荐系统 TransE向量表示	系统审计日志 (含进程/文件/套接字/调用关系)	APT攻击	通过推荐概念和知识图谱来识别网络威胁	√	√	●	√	√	M	√

表 6 APT 攻击智能推理工作总结表 (续)

分类	相关工作	推理方法	推理特征	应用场景	任务及特点	关系图谱	未知推理	先验知识	真实场景	DNN	鲁棒性	实时检测
攻击意图推理	WATSON[60]	知识图谱+语义推理 HCA聚类算法	系统审计日志 (进程/文件/关系)	APT攻击	通过语义推理来表征行为	√	√	●	√	×	M	√
	ATLAS[176]	知识图谱+溯源图 LSTM+序列 因果关系图	系统审计日志 (含进程/文件/套接字/调用关系)	APT攻击	生成抽象攻击策略和构建APT攻击故事	√	×	●	×	√	L	×
攻击路径感知	陈伟翔[180]	隐马尔可夫模型 关键基因序列 恶意行为序列	APT恶意软件 基因序列特征 恶意行为	APT攻击	根据APT恶意行为链来预测后续攻击路径	√	√	●	√	×	L	×
	胡浩等人[185]	吸收Markov链 攻击图	模拟DDoS攻击(IP/请求/漏洞等)	网络渗透	构建模型预测网络入侵路径	√	×	●	×	×	L	×
	CONAN[35]	有限状态自动机 状态检测框架 智能策略	审计日志 (含进程/文件/套接字/API调用)	APT攻击	利用状态关系还原攻击链和检查APT攻击	√	×	●	√	×	M	√
	Patrol[190]	依赖关系图 搜索算法 规则匹配	系统调用(含进程/文件/套接字/关系)	0day漏洞	利用系统对象依赖关系推理0day攻击路径	√	×	●	√	×	L	√
	Kim等人[193]	基于网络的推理 定制信念传播算法 边缘分配机制	网络钓鱼URL相关特征 (含URL/域名/IP等)	网络钓鱼	基于网络推理来检测逃逸型网络钓鱼URL	√	√	●	√	×	M	√
攻击场景还原	伏晓等人[200]	告警关联+依赖追踪 层次化映射	IDS报警和操作系统层时间记录	入侵检测	高层行为场景和低层行为细节重构	√	×	●	√	×	L	×
	SLEUTH[34]	溯源图+依赖关系图 定制策略和属性	系统审计日志 (进程/文件/关系)	APT攻击	首次利用溯源图重构场景	√	√	●	√	×	L	√
	MORSE[204]	溯源图+依赖关系图 标签衰减+路径优化	系统审计日志 (进程/文件/关系)	APT攻击	减轻依赖爆炸以构建场景图	√	√	●	√	×	H	√
	HOLMES[51]	溯源图+杀伤链模型 高级场景图 依赖关系剪枝	系统审计日志 (进程/文件/关系) 攻击链场景知识	APT攻击	实时检测APT攻击并消除语义鸿沟	√	√	●	√	×	H	√
	Extrator[209]	溯源图 自然语言处理 语义分析	系统审计日志 (进程/文件/关系) 威胁情报知识	APT攻击	利用NLP自动提取威胁情报并生成溯源图	√	√	◑	√	√	M	√
攻击阻断及反制	Foureye[53]	超博弈理论 防御欺骗技术	模拟攻击流量 (蜜罐/设备/攻击)	APT攻击	感知攻击各阶段并实现防御	×	×	●	×	×	L	×
	Dai[230]	启发式蜜罐 Snort+OpenFlow ATT&CK框架	APT组织WebShell和流量	APT攻击	构建多平台蜜罐来识别和阻断APT攻击	×	√	◑	√	×	L	×
	Velucha-my[231]	深度强化学习 蜜罐	物联网数据集 DoS攻击流量	DoS攻击	实时检测和阻断DoS攻击	×	√	◑	√	√	M	√

注: ●表示需要大量先验知识或未考虑过度依赖先验知识的问题; ◑表示需要少量先验知识或较少依赖先验知识; ○表示需要极少或不需要先验知识; L (low)表示鲁棒性较差; M (middle)表示鲁棒性中等或具有一定迁移性; H (high)表示鲁棒性好; √表示利用该方法或涉及该内容; ×表示未利用该方法或未涉及该内容.

4.5 推理对抗

先前章节详细综述和分析 APT 攻击智能推理 4 类关键任务的相关工作, 也介绍 4 类任务智能方法的特点和面临的挑战, 对抗和躲避攻击推理和检测的趋势越发明显. 现阶段针对 APT 攻击的推理对抗主要分为 3 类.

(1) 利用混淆、加壳、加密、API 钩子和秘密通信技术来躲避检测系统的推理. 攻击者会根据检测系统静态分析和动态分析的特点, 对恶意软件进行加壳、混淆处理来防止提取静态特征和识别恶意行为, 同时利用 API 钩子等技术来防止动态识别关键 API 序列和功能函数. 随后, 解混淆、加壳识别、动态行为提取技术被提出来防御这类传统的推理及检测对抗技术. Li 等人[44]详细分析 PowerShell 代码混淆技术, 提出有效的解混淆方法. Cheng

等人[240]系统研究 Windows 恶意软件加壳技术, 并提出一种名叫 BinUnpack 的脱壳方法, 能有效实现大规模 Windows 恶意软件的脱壳.

(2) 构建融合反调试 (anti-debugging)、反虚拟化 (anti-virtualization)、反沙箱 (anti-sandbox)、反分析 (anti-analysis)、反仿真 (anti-emulation)、环境克隆和代码重用技术的恶意软件来对抗推理[135,241]. 尽管产业界会利用虚拟化技术来监控恶意软件行为以保护底层操作系统, 然而, APT 组织或攻击者会利用该类对抗技术来躲避入侵检测系统或病毒分析引擎的分析和推理, 甚至通过已有且常用的 API、组件或应用程序实施高隐蔽攻击并对抗检测[242,243]. Galloro 等人[244]详细分析恶意软件作者用来隐藏恶意活动并阻碍安全分析的规避行为, 系统收集并归纳了 Windows 恶意软件用来对抗检测环境 (如调试器和虚拟机) 的 92 种规避技术, 并开发一个细粒度框架来识别真实恶意样本的规避行为信息. 该类对抗可以通过用户行为模拟[245]、知识框架及特征增强[241]和关联分析来避免.

(3) 通过对抗样本和数据投毒攻击来欺骗智能检测模型, 窃取机器学习或深度学习模型的参数及隐私. 近年来, 随着人工智能技术被广泛应用于网络安全、计算机视觉、自动驾驶等领域, 该技术正在改变人们的生活方式与促进社会进步. 然而, 智能模型的应用与部署存在严峻的安全缺陷. 攻击者会围绕深度学习模型的完整性、鲁棒性和隐私性开展攻击, 随之产生了对应的对抗样本攻击、推断攻击、投毒攻击、模型窃取攻击[246]. 攻击者会在恶意样本中添加噪声或对抗特征, 干扰深度学习模型训练, 模型的预测结果会发生偏移, APT 攻击检测模型错误地将恶意样本识别为良性. 此外, 攻击者会针对特定数据集或环境, 结合智能模型预测结果推断和窃取深度学习模型的数据和敏感信息, 甚至还原恶意检测模型. 针对该类对抗, 学术界和产业界从硬件和软件安全保护两个角度开展研究, 文献 [247] 是经典的神经网络后门攻击识别与缓解方法, 通过神经元剪枝、滤波器和遗忘学习来缓解.

4.6 小 结

本节主要介绍面向 APT 攻击的智能推理研究, 结合 APT 攻击智能推理框架分别归纳和综述 4 个代表性子任务, 包括攻击意图推理、攻击路径感知、攻击场景还原和攻击阻断及反制, 通过各子任务的相互关联和递进, 能更好地实现 APT 攻击防御. 整体而言, APT 攻击智能推理仍处于起步阶段, APT 组织具有较强的专业性, 使用的技术手段和方案较为先进和成熟, 甚至会采用代码混淆、模拟欺骗等手段对抗检测、溯源和推理系统, APT 攻击的推理和反制存在诸多难点. 未来, 结合攻击检测、溯源和推理的整体防御链将更加完善, 利用人工智能、知识图谱并融合 APT 攻击全生命周期的推理技术将成为新的热点, 通过攻击意图推理、路径感知、场景还原能更好地阻断和反制 APT 攻击, 并且学术界和产业界将取得新的突破, 弥补整个领域存在的不足.

5 研究趋势讨论与分析

本文围绕 APT 攻击特点提出 APT 攻击防御链, 对 APT 攻击智能检测、溯源和推理 3 个方面的工作进行详细总结和阐述. 基于前文的综述, 该部分将探讨和分析 APT 攻击的热点主题关联及未来的研究趋势.

5.1 APT 攻击领域的研究热点主题及关联

首先, 结合 APT 攻击检测、溯源和推理相关的 200 多篇会议与期刊论文 (大部分为网络与信息安全领域的 CCF 系列论文), 对论文的关键词及核心主题进行提取, 经过预处理并提取实体和关系后, 形成如图 14 所示的 APT 攻击领域的主题关系图谱. 由图 4 可知, 整个领域划分为 4 个主题群, 具体内容如下.

(1) 位于中心 A1 区域的是 APT 攻击和人工智能技术主题群, 典型的热点主题包括“advanced persistent threat”“APT”“deep learning”“machine learning”“cyber attack”“APT malware”等, 即符合本文研究的主题;

(2) 位于左下角 A2 区域的是 APT 攻击检测主题群, 热点主题包括“malware”“detection”“classification”“APT detection”“malware behavior”等, 涵盖恶意性检测、恶意家族检测、恶意行为识别和恶意代码定位任务;

(3) 位于最右边 A3 区域的是 APT 攻击溯源主题群, 热点主题包括“attribution”“traceback”“APT attribution”“de-anonymizing”“ip traceback”“authorship attribution”等, 涵盖多种目标的溯源主题;

(4) 位于左上角 A4 区域的是 APT 攻击推理主题群, 热点主题包括“reasoning”“knowledge reasoning”“knowledge graph”“ontology”“threat analysis”等, 基于知识图谱的攻击推理和感知是当前研究热点.

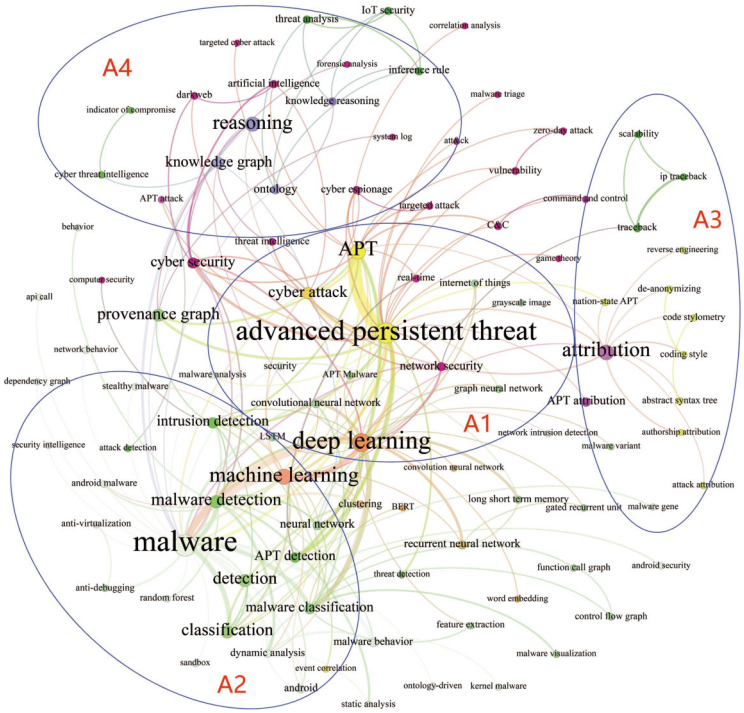

图 14 APT 攻击领域热点主题关系图谱

整体而言, APT 攻击检测、溯源和推理领域围绕着 APT 攻击和智能技术开展, 各主题之间紧密联系又存在差异, 整个研究趋势从检测到溯源, 再到推理逐步发展, 并且智能程度逐渐提升. 究其原因, 一方面面对日益增加的海量恶意软件和 APT 攻击事件, 智能化的方法能够迅速、准确地识别大规模的攻击行为, 挖掘深层次的语义知识和依赖关系, 并溯源攻击来源, 这些提升均是传统方法无法实现的; 另一方面, 网络攻防是博弈的过程, APT 攻击者会利用加壳、混淆、对抗和逃逸等手段来躲避检测及溯源, 而传统方法过度依赖专家经验, 需要持续地添加新规则和特征, 导致无法检测未知且变化的 APT 攻击, 而智能方法旨在弥补现有检测系统的不足, 通过学习和关联来发现未知的攻击, 识别细粒度行为, 甚至推理后续攻击意图. 总之, 尽管现有研究已取得一定成果, 但是面对日益变化, 隐蔽性、逃逸性和对抗性逐渐增强的攻击, 现有方法仍然存在诸多挑战, 本文将在第 5.2 节概述.

5.2 APT 攻击领域的研究趋势分析

随着全球网络安全形式日益复杂, 如何快速精准地检测、溯源和推理 APT 攻击已经成为重要的研究主题, 智能化 APT 攻击防护能有效保障国家及全民的安全, 更好地检测攻击行为、溯源攻击来源和推理攻击意图, 并帮助安全分析人员及时掌握网络安全态势. 针对先前的综述和分析, 结合 APT 攻击智能检测、溯源和推理 3 部分研究内容的缺点和挑战, 该部分将对整个 APT 领域未来的研究趋势进行展望. 具体内容如下.

(1) 融合 APT 攻击全生命周期的防御机制亟需加强, 结合检测、溯源和推理开展 APT 攻击防御研究

APT 攻击具有较高的组织性和持续性, 传统防御方法忽视了 APT 攻击的生命周期, 并将检测、溯源、推理分割开来研究, 而这些高隐蔽攻击各个阶段通常存在紧密联系, 前后的信息传递也存在关联. 随着网络杀伤链、ATT&CK 等框架被提出, 学术界和产业界越来越重视对 APT 全生命周期的研究, 并且通过将各阶段特征和数据转换为知识和关系, 能更好地发现攻击、追踪来源, 甚至推理攻击意图并阻断攻击. 基于此, 本文预测未来将出现

更多融合 APT 攻击全生命周期的防御方法, 并且贯穿攻击检测、溯源和推理.

(2) 结合主动防御和前置防御将 APT 攻击提前阻断或从源头上挫败恶意网络活动

相较于美国在网络战略中提出防御前置 (defend forward) 的战略构想[248,249], 其旨在构建主动防御和前置防御技术从源头上阻断和挫败恶意网络活动以及 APT 攻击, 本文工作的重点主要集中于被动防御技术, 并详细阐述智能化技术应用于 APT 攻击检测、溯源及推理研究. 因此, 本文工作存在一定的局限性且无法及时构建完整的证据链, 尤其难以满足当代网络环境下对 APT 攻击的实时精准溯源以及将攻击阻断在源头的需求. 基于此, 结合美国所提出的前置防御技术, 本文预测未来的网络空间安全及 APT 防御会逐渐从被动防御转向主动防御, 会力争将网络攻击及 APT 行动从源头上进行干预和阻断, 甚至结合蜜罐技术采用先发制人的方式来遏制网络威胁, 从而维护国家及人民的利益. 此外, 主动防御及前置防御技术也是对网络攻击链及防御链的完善和补充, 结合前置防御的 APT 攻击检测、溯源及推理能更好地实施 APT 防御, 以至于构建更有效的防御体系.

(3) AI 赋能的 APT 攻击自动化和智能化防御技术将是未来研究的重点

当前 APT 攻击防御面临的一个巨大挑战是无法准确地将不同形式的网络攻击事件关联起来, 缺乏对 APT 攻击组织画像的细粒度刻画, 更无法构建 APT 攻击全貌和推理深层次的攻击意图. 究其原因是 APT 攻击检测和溯源仍过度依赖人工经验和专家知识, 较少有团队和机构能对 APT 攻击开展大规模的研究, 传统检测方法无法有效对抗混淆、伪装、逃逸和欺骗性强的 APT 攻击. 此外, 随着大语言模型和 ChatGPT 技术火热发展, 本文预测未来将逐渐出现利用相关技术生成对抗样本、实现隐蔽信息传输、智能问答式社会工程学攻击以及模拟伪装成不同 APT 组织的攻击事件, 从而严重影响全球的网络安全. 因此, 如何利用人工智能技术阻止该类攻击发生, 如何准确提取 APT 攻击的代码基因和语义知识, 有效识别伪装的攻击和阻断对抗样本攻击, 以及理解攻击意图, 实现更加智能化、自动化和无监督式的整体防御将成为未来研究的重点.

(4) 面向 Lotl 技术的高隐蔽 APT 攻击防御需取得突破

APT 组织更倾向于利用 Lotl 技术开展离地或无文件攻击. 这类方法更少地留存痕迹在被攻击系统中, 传统入侵检测系统或安全防火墙无法捕获攻击特征及系统日志. 在此背景下, 如何更好地构建智能化方法提取动态运行时的状态和信息, 有效识别恶意攻击将变得更加重要[250]. 总之, 面向 Lotl 技术的 APT 攻击仍存在诸多难点, 这将是未来亟需突破的重要研究任务.

(5) 基于 0day 漏洞的 APT 攻击防御需进一步加强

APT 组织会通过多个 0day 漏洞渗透内部网络、窃取重要情报、破坏关键基础设施, 甚至基于上游供应链软件或可信系统实施攻击. 因此, 通过构建不同设备或环境的知识库, 利用漏洞自动扫描算法分析不同行为产生的前后差异, 有效识别系统或网络的漏洞也变得至关重要, 这将有效保障关键基础设施和物联网设备的安全.

(6) 面向对抗性攻击、环境逃逸和概念漂移的 APT 攻击防御需深化, 并且结合底层硬件交互开展防护

随着对抗样本和投毒攻击技术的发展, 传统恶意代码检测模型会被误导, 甚至其模型参数和隐私会被窃取. 同时, 针对环境、沙箱、云端逃逸的研究逐渐增加, 攻击者会构建变体程序来实施网络攻击. 如果仅根据攻击所产生的特征信息或日志数据, 是无法有效溯源和推理该类 APT 攻击. 因此, 通过底层硬件 (如编译器和汇编代码) 数据交互来开展 APT 攻击防御将是未来研究的重点, 如何克服概念漂移和数据漂移也是智能模型需要考虑的关键因素.

(7) 通过挖掘高层次语义知识来检测、溯源和推理新型和未知的 APT 攻击, 提高模型的可解释性, 进一步将智能模型应用于实际安全场景及产品中

低层次攻击数据和高层次攻击行为存在巨大的语义鸿沟, 如何挖掘和利用高层次语义知识存在诸多难点, 在新型和未知的 APT 攻击事件中更甚. 因此, 如何构建智能模型和轻量级、语义丰富的溯源图是该领域研究的重点和难点. 同时, 由于深度学习模型的可解释性较差, 如何利用知识图谱来刻画 APT 攻击知识将是另一个研究趋势, 通过事件上下文关联和攻击依赖关系分析, 能增加模型的可解释性, 并且能更好地推理和发现后续的攻击行为、

意图和攻击路径, 从而及时阻断和反制攻击. 最后, 产业界和学术界均倾向于将智能模型应用于实际安全场景和安全产品中, 最终推动整个领域的发展, 为真实的世界提供更好的安全防护.

综上所述, 该部分结合 APT 攻击特点、现状和先前的综述研究, 归纳总结了 APT 攻击领域未来的 6 大研究趋势, 它们将力争更好地保障全球网络空间安全.

6 总 结

本文详细总结当前 APT 攻击智能防护相关工作的研究概况, 提出了 APT 攻击防御链, 分别对 APT 攻击检测、APT 攻击溯源和 APT 攻击推理的智能化方法进行系统、深入、全面地总结和关联. 首先, 详细分析面向区域、组织、攻击者、地址和攻击模型的 APT 溯源研究; 其次, 从攻击意图推理、攻击路径感知、攻击场景还原、攻击阻断和反制 4 个方面详细归纳和对比 APT 攻击智能推理研究; 最后, 对 APT 攻击防御领域的热点主题、发展趋势和挑战进行讨论.

此外, 本文通过系统性总结和分析, 尝试回答了 APT 攻击领域的 4 类研究问题: ① 在第 1.2 节给出 APT 攻击防御链, 并有效将 APT 攻击检测、溯源和推理系统关联; ② 分别在第 2 节和第 3 节详细梳理和归纳 APT 攻击中溯源和推理的现有方法特点, 并概述不同方法和策略的基本思路及优缺点; ③ 智能化方法通过挖掘深层语义知识和特征, 构建安全知识图谱或溯源图来提升对高隐蔽、高对抗、高威胁的 APT 攻击检测; ④ 在第 5.2 节对 APT 攻击未来的研究趋势及挑战进行阐述.

总之, 本文从多个维度综述 APT 攻击的智能防御方法, 结合前沿和智能视角分析 APT 攻击研究现状, 并将检测、溯源和推理相结合来构建 APT 攻击整体防御, 并展望了该研究未来的研究方向. 希望本文能为今后 APT 攻击防御和网络安全及相关研究提供帮助, 进一步推动 APT 攻击智能检测、溯源和推理的发展.

References:

[1] Tankard C. Advanced persistent threats and how to monitor and deter them. Network Security, 2011, 2011(8): 16–19. [doi: 10.1016/S1353-4858(11)70086-1]

[2] Binde BE, McRee R, O'Connor TJ. Assessing outbound traffic to uncover advanced persistent threat. Technical Report, SANS Technology Institute, 2011. [doi: 10.13140/RG.2.2.16401.07520]

[3] Fu Y, Li HC, Wu XP, Wang JS. Detecting APT attacks: A survey from the perspective of big data analysis. Journal on Communications, 2015, 36(11): 1–14 (in Chinese with English abstract). [doi: 10.11959/j.issn.1000-436x.2015184]

[4] Stojanović B, Hofer-Schmitz K, Kleb U. APT datasets and attack modeling for automated detection methods: A review. Computers & Security, 2020, 92: 101734. [doi: 10.1016/j.cose.2020.101734]

[5] Chen RD, Zhang XS, Niu WN, Lan HY. A research on architecture of APT attack detection and countering technology. Journal of University of Electronic Science and Technology of China, 2019, 48(6): 870–879 (in Chinese with English abstract). [doi: 10.3969/j.issn.1001-0548.2019.06.011]

[6] Langner R. Stuxnet: Dissecting a cyberwarfare weapon. IEEE Security & Privacy, 2011, 9(3): 49–51. [doi: 10.1109/MSP.2011.67]

[7] Antiy Laboratory. A comprehensive analysis report on Ukraine power grid outage. 2016 (in Chinese). https://www.antiy.com/response/A_Comprehensive_Analysis_Report_on_Ukraine_Power_Grid_Outage/A_Comprehensive_Analysis_Report_on_Ukraine_Power_Grid_Outage.html

[8] FireEye. Highly evasive attacker leverages SolarWinds supply chain to compromise multiple global victims with SUNBURST backdoor. 2020. https://www.fireeye.com/blog/threat-research/2020/12/evasive-attacker-leverages-solarwinds-supply-chain-compromises-with-sunburst-backdoor.html

[9] Digital Security Technology Group. Investigative report on northwestern Polytechnical University's discovery of US NSA cyber attack. 2022 (in Chinese). https://mp.weixin.qq.com/s/0ReOzQMM5GS4xXRUPpKCvA

[10] Alshamrani A, Myneni S, Chowdhary A, Huang DJ. A survey on advanced persistent threats: Techniques, solutions, challenges, and research opportunities. IEEE Communications Surveys & Tutorials, 2019, 21(2): 1851–1877. [doi: 10.1109/COMST.2019.2891891]

[11] Zhang Y, Pan XM, Liu QZ, Cao JK, Luo ZQ. APT attacks and defenses. Journal of Tsinghua University (Science & Technology), 2017, 57(11): 1127–1133 (in Chinese with English abstract). [doi: 10.16511/j.cnki.qhdxxb.2017.21.024]

[12] Ye YF, Li T, Adjeroh D, Iyengar SS. A survey on malware detection using data mining techniques. ACM Computing Surveys, 2018,

50(3): 41. [doi: 10.1145/3073559]

[13] Wang R, Feng DG, Yang Y, Su PR. Semantics-based malware behavior signature extraction and detection method. Ruan Jian Xue Bao/Journal of Software, 2012, 23(2): 378–393 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3953.htm [doi: 10.3724/SP.J.1001.2012.03953]

[14] Song WN, Peng GJ, Fu JM, Zhang HG, Chen SL. Research on malicious code evolution and traceability technology. Ruan Jian Xue Bao/Journal of Software, 2019, 30(8): 2229–2267 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5767.htm [doi: 10.13328/j.cnki.jos.005767]

[15] Yang XZ, Peng GJ, Luo Y, Song WN, Zhang J, Cao FT. OMRDetector: A method for detecting obfuscated malicious requests based on deep learning. Chinese Journal of Computers, 2022, 45(10): 2167–2189 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2022.02167]

[16] Ongun T, Stokes J W, Or JB, Tian K, Tajaddodianfar F, Neil J, Seifert C, Oprea A, Platt JC. Living-off-the-land command detection using active learning. In: Proc. of the 24th Int'l Symp. on Research in Attacks, Intrusions and Defenses. San Sebastian: ACM, 2021. 442–455. [doi: 10.1145/3471621.3471858]

[17] Wang Q, Hassan WU, Li D, Jee K, Yu X, Zou KX, Rhee J, Chen ZZ, Cheng W, Gunter CA, Chen HF. You are what you do: Hunting stealthy malware via data provenance analysis. In: Proc. of the 27th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2020. [doi: 10.14722/ndss.2020.24167]

[18] Faghihi F, Zulkernine M, Ding S. CamoDroid: An Android application analysis environment resilient against sandbox evasion. Journal of Systems Architecture, 2022, 125: 102452. [doi: 10.1016/j.sysarc.2022.102452]

[19] Nasr M, Bahramali A, Houmansadr A. DeepCorr: Strong flow correlation attacks on tor using deep learning. In: Proc. of the 2018 ACM SIGSAC Conf. on Computer and Communications Security. Toronto: ACM, 2018. 1962–1976. [doi: 10.1145/3243734.3243824]

[20] Crandall JR, Su ZD, Wu SF, Chong FT. On deriving unknown vulnerabilities from zero-day polymorphic and metamorphic worm exploits. In: Proc. of the 12th ACM Conf. on Computer and Communications Security. Alexandria: ACM, 2005. 235–248. [doi: 10.1145/1102120.1102152]

[21] Guan SP, Jin XL, Jia YT, Wang YZ, Cheng XQ. Knowledge reasoning over knowledge graph: A survey. Ruan Jian Xue Bao/Journal of Software, 2018, 29(10): 2966–2994 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5551.htm [doi: 10.13328/j.cnki.jos.005551]

[22] Chou D, Jiang M. A survey on data-driven network intrusion detection. ACM Computing Surveys, 2022, 54(9): 182. [doi: 10.1145/3472753]

[23] Ma YX, Zhang QX, Tan YA, Shen M. Research on Behavior-prediction for Intelligent Attacks. Ruan Jian Xue Bao/Journal of Software, 2021, 32(5): 1526–1546 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6204.htm [doi: 10.13328/j.cnki.jos.006204]

[24] Luo Y, Xiao Y, Cheng L, Peng GJ, Yao DF. Deep learning-based anomaly detection in cyber-physical systems: Progress and opportunities. ACM Computing Surveys, 2022, 54(5): 106. [doi: 10.1145/3453155]

[25] Liu B, Ding M, Shaham S, Rahayu W, Farokhi F, Lin ZH. When machine learning meets privacy: A survey and outlook. ACM Computing Surveys, 2022, 54(2): 31. [doi: 10.1145/3436755]

[26] Liras LFM, de Soto AR, Prada MA. Feature analysis for data-driven APT-related malware discrimination. Computers & Security, 2021, 104: 102202. [doi: 10.1016/j.cose.2021.102202]

[27] Downing E, Mirsky Y, Park K, Lee W. DeepReflect: Discovering malicious functionality through binary reconstruction. In: Proc. of the 30th USENIX Security Symp. Vancouver: USENIX Association, 2021. 3469–3486.

[28] Li ZT, Cheng X, Sun LX, Zhang J, Chen B. A hierarchical approach for advanced persistent threat detection with attention-based graph neural networks. Security and Communication Networks, 2021, 2021: 9961342. [doi: 10.1155/2021/9961342]

[29] Han XY, Pasquier TFJM, Bates A, Mickens J, Seltzer MI. Unicorn: Runtime provenance-based detector for advanced persistent threats. In: Proc. of the 27th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2020.

[30] Berninger M. Going ATOMIC: Clustering and associating attacker activity at scale. 2011. https://www.mandiant.com/resources/clustering-and-associating-attacker-activity-at-scale

[31] Alrabaee S, Shirani P, Debbabi M, Wang LY. On the feasibility of malware authorship attribution. In: Proc. of the 9th Foundations and Practice of Security. Québec City: Springer, 2016. 256–272. [doi: 10.1007/978-3-319-51966-1_17]

[32] Li A. Research on features extraction method of attack group based on malicious code gene [MS. Thesis]. Beijing: Beijing University of Posts and Telecommunications, 2021 (in Chinese with English abstract). [doi: 10.26969/d.cnki.gbydu.2021.001346]

[33] Huang KZ, Lian YF, Feng DG, Zhang HX, Wu D, Ma XL. Method of cyber attack attribution based on graph model. Ruan Jian Xue Bao/Journal of Software, 2022, 33(2): 683–698 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6314.htm [doi: 10.

13328/j.cnki.jos.006314]

[34] Hossain N, Milajerdi SM, Wang JA, Eshete B, Gjomemo R, Sekar R, Stoller SD, Venkatakrishnan VN. SLEUTH: Real-time attack scenario reconstruction from COTS audit data. In: Proc. of the 26th USENIX Conf. on Security Symp. Vancouver: USENIX Association, 2017. 487–504.

[35] Xiong CL, Zhu TT, Dong WH, Ruan LQ, Yang RQ, Cheng YQ, Chen Y, Cheng S, Chen XT. Conan: A practical real-time APT detection system with high accuracy and efficiency. IEEE Trans. on Dependable and Secure Computing, 2022, 19(1): 551–565. [doi: 10.1109/TDSC.2020.2971484]

[36] Zimba A, Chen HS, Wang ZS. Bayesian network based weighted APT attack paths modeling in cloud computing. Future Generation Computer Systems, 2019, 96: 525–537. [doi: 10.1016/j.future.2019.02.045]

[37] Lemay A, Calvet J, Menet F, Fernandez JM. Survey of publicly available reports on advanced persistent threat actors. Computers & Security, 2018, 72: 26–59. [doi: 10.1016/j.cose.2017.08.005]

[38] Talib MA, Nasir Q, Nassif AB, Mokhamed T, Ahmed N, Mahfood B. APT beaconing detection: A systematic review. Computers & Security, 2022, 122: 102875. [doi: 10.1016/j.cose.2022.102875]

[39] Liu CG, Fang BX, Liu BX, Cui X, Liu QX. A hierarchical model of targeted cyber attacks attribution. Journal of Cyber Security, 2019, 4(4): 1–18 (in Chinese with English abstract). [doi: 10.19363/J.cnki.cn10-1380/tn.2019.07.01]

[40] Pan YF, Zhu JH, Zhou TY. Survey on APT attack scenario reconstruction methods. Journal of Information Engineering University, 2021, 22(1): 55–60, 80 (in Chinese with English abstract). [doi: 10.3969/j.issn.1671-0673.2021.01.010]

[41] Chen ZY, Liu JX, Shen Y, Simsek M, Kantarci B, Mouftah HT, Djukic P. Machine learning-enabled IoT security: Open issues and challenges under advanced persistent threats. ACM Computing Surveys, 2023, 55(5): 105. [doi: 10.1145/3530812]

[42] 360 CERT. Global advanced persistent threat (APT) research report for the first half of 2021. 2021 (in Chinese). https://cert.360.cn/report/detail?id=6c9a1b56e4ceb84a8ab9e96044429adc

[43] MITRE Corporation. Groups MITRE ATT&CK. 2022. https://attack.mitre.org/groups/

[44] Li ZY, Chen QA, Xiong CL, Chen Y, Zhu TT, Yang H. Effective and light-weight deobfuscation and semantic-aware attack detection for PowerShell scripts. In: Proc. of the 2019 ACM SIGSAC Conf. on Computer and Communications Securit. London: ACM, 2019. 1831–1847. [doi: 10.1145/3319535.3363187]

[45] FireEye. M-Trends 2021: FireEye mandiant services. 2021. https://www.arrow.com/ecs-media/16352/fireeye-rpt-mtrends-2021.pdf

[46] Martin L. Cyber kill chain. 2022. https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

[47] Oosthoek K, Doerr C. SoK: ATT&CK techniques and trends in Windows malware. In: Proc. of the 15th Int'l Conf. on Security and Privacy in Communication Systems. Orlando: Springer, 2019. 406–425. [doi: 10.1007/978-3-030-37228-6_20]

[48] 360 Beaconlab. Analysis of Darkhotel (APT-C-06) using "Double Star" 0day vulnerability (CVE-2019–17026, CVE-2020–0674) to launch APT attack to China. 2020 (in Chinese). http://pub1-bjyt.s3.360.cn/bcms/Darkhotel%EF%BC%88APT-C-06%EF%BC%89%E4%BD%BF%E7%94%A8%E2%80%9C%E5%8F%8C%E6%98%9F%E2%80%9D0Day%E6%BC%8F%E6%B4%9E%EF%BC%88CVE-2019-17026%E3%80%81CVE-2020-0674%EF%BC%89%E9%92%88%E5%AF%B9%E4%B8%AD%E5%9B%BD%E5%8F%91%E8%B5%B7%E7%9A%84APT%E6%94%BB%E5%87%BB%E5%88%86%E6%9E%90.pdf

[49] Wang JL, Zhang C, Qi XY, Rong Y. A survey of intelligent malware detection on windows platform. Journal of Computer Research and Development, 2021, 58(5): 977–994 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2021.20200964]

[50] Yoo S, Kim S, Kim S, Kang BB. AI-HydRa: Advanced hybrid approach using random forest and deep learning for malware classification. Information Sciences, 2021, 546: 420–435. [doi: 10.1016/j.ins.2020.08.082]

[51] Milajerdi SM, Gjomemo R, Eshete B, Sekar R, Venkatakrishnan VN. HOLMES: Real-time APT detection through correlation of suspicious information flows. In: Proc. of the 2019 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2019. 1137–1152. [doi: 10.1109/SP.2019.00026]

[52] Milajerdi SM, Eshete B, Gjomemo R, Venkatakrishnan VN. POIROT: Aligning attack behavior with kernel audit records for cyber threat hunting. In: Proc. of the 2019 ACM SIGSAC Conf. on Computer and Communications Security. London: ACM, 2019. 1795–1812. [doi: 10.1145/3319535.3363217]

[53] Wan ZL, Cho JH, Zhu M, Anwar AH, Kamhoua CA, Singh MP. Foureye: Defensive deception against advanced persistent threats via hypergame theory. IEEE Trans. on Network and Service Management, 2022, 19(1): 112–129. [doi: 10.1109/TNSM.2021.3117698]

[54] Bolton AD, Anderson-Cook CM. APT malware static trace analysis through bigrams and graph edit distance. Statistical Analysis and Data Mining: The ASA Data Science Journal, 2017, 10(3): 182–193. [doi: 10.1002/sam.11346]

[55] Dahl GE, Stokes JW, Deng L, Yu D. Large-scale malware classification using random projections and neural networks. In: Proc. of the 2013 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. Vancouver: IEEE, 2013. 3422–3426. [doi: 10.1109/ICASSP.2013.6638293]

[56] Athiwaratkun B, Stokes JW. malware classification with LSTM and GRU language models and a character-level CNN. In: Proc. of the 2017 IEEE Int'l Conf. on Acoustics, Speech and Signal Processing. New Orleans: IEEE, 2017. 2482–2486. [doi: 10.1109/ICASSP.2017.7952603]

[57] Demirkıran F, Çayır A, Ünal U, Dağ H. An ensemble of pre-trained Transformer models for imbalanced multiclass malware classification. Computers & Security, 2022, 121: 102846. [doi: 10.1016/j.cose.2022.102846]

[58] Çayır A, Ünal U, Dağ H. Random CapsNet forest model for imbalanced malware type classification task. Computers & Security, 2021, 102: 102133. [doi: 10.1016/j.cose.2020.102133]

[59] Ding YX, Dai W, Yan SL, Zhang YM. Control flow-based opcode behavior analysis for malware detection. Computers & Security, 2014, 44: 65–74. [doi: 10.1016/j.cose.2014.04.003]

[60] Zeng J, Chua ZL, Chen YF, Ji KH, Liang ZK, Mao J. WATSON: Abstracting behaviors from audit logs via aggregation of contextual semantics. In: Proc. of the 28th Annual Network and Distributed System Security Symp. The Internet Society, 2021. [doi: 10.14722/ndss.2021.24549]

[61] Alrawi O, Ike M, Pruett M, Kasturi RP, Barua S, Hirani T, Hill B, Saltaformaggio B. Forecasting malware capabilities from cyber attack memory images. In: Proc. of the 30th USENIX Security Symp. USENIX Association, 2021. 3523–3540.

[62] Xiao D, Liu BH, Cui BJ, Wang XC, Zhang SX. Malware prediction technique based on program gene. Chinese Journal of Network and Information Security, 2018, 4(8): 21–30 (in Chinese with English abstract). [doi: 10.11959/j.issn.2096-109x.2018069]

[63] Mirzaei O, Vasilenko R, Kirda E, Lu L, Kharraz A. SCRUTINIZER: Detecting code reuse in malware via decompilation and machine learning. In: Proc. of the 18th Int'l Conf. on Detection of Intrusions and Malware, and Vulnerability Assessment. Springer, 2021. 130–150. [doi: 10.1007/978-3-030-80825-9_7]

[64] Ghafir I, Hammoudeh M, Prenosil V, Han LX, Hegarty R, Rabie K, Aparicio-Navarro FJ. Detection of advanced persistent threat using machine-learning correlation analysis. Future Generation Computer Systems, 2018, 89: 349–359. [doi: 10.1016/j.future.2018.06.055]

[65] Mamun M, Shi K. DeepTaskAPT: Insider APT detection using task-tree based deep learning. In: Proc. of the 20th IEEE Int'l Conf. on Trust, Security and Privacy in Computing and Communications. Shenyang: IEEE, 2021. 693–700. [doi: 10.1109/TrustCom53373.2021.00102]

[66] Gao Y, Hasegawa H, Yamaguchi Y, Shimada H. malware detection using attributed CFG generated by pre-trained language model with graph isomorphism network. In: Proc. of the 46th Annual Computers, Software, and Applications Conf. Los Alamitos: IEEE, 2022. 1495–1501. [doi: 10.1109/COMPSAC54236.2022.00237]

[67] Yu ZP, Cao R, Tang QY, Nie S, Huang JZ, Wu S. Order matters: Semantic-aware neural networks for binary code similarity detection. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. New York: AAAI Press, 2020. 1145–1152. [doi: 10.1609/aaai.v34i01.5466]

[68] Chanajitt R, Pfahringer B, Gomes HM. Combining static and dynamic analysis to improve machine learning-based malware classification. In: Proc. of the 8th IEEE Int'l Conf. on Data Science and Advanced Analytics. Porto: IEEE, 2021. 1–10. [doi: 10.1109/DSAA53316.2021.9564144]

[69] Gibert D, Mateu C, Planes J, Vicens R. Classification of malware by using structural entropy on convolutional neural networks. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence. New Orleans: AAAI Press, 2018. 7759–7764. [doi: 10.1609/aaai.v32i1.11409]

[70] Herath JD, Wakodikar PP, Yang P, Yan GH. CFGExplainer: Explaining graph neural network-based malware classification from control flow graphs. In: Proc. of the 52nd Annual IEEE/IFIP Int'l Conf. on Dependable Systems and Networks. Baltimore: IEEE, 2022. 172–184. [doi: 10.1109/DSN53405.2022.00028]

[71] Gao XW, Hu CZ, Shan C, Han WJ. MaliCage: A packed malware family classification framework based on DNN and GAN. Journal of Information Security and Applications, 2022, 68: 103267. [doi: 10.1016/j.jisa.2022.103267]

[72] Das S, Xiao H, Liu Y, Zhang W. Online malware defense using attack behavior model. In: Proc. of the 2016 IEEE Int'l Symp. on Circuits and Systems. Montréal: IEEE, 2016. 1322–1325. [doi: 10.1109/ISCAS.2016.7527492]

[73] Zhang H, Zhang WJ, Lv ZH, Sangaiah AK, Huang T, Chilamkurti N. MALDC: A depth detection method for malware based on behavior chains. World Wide Web, 2020, 23(2): 991–1010. [doi: 10.1007/s11280-019-00675-z]

[74] Xuan CT, Copeland J, Beyah R. Toward revealing kernel malware behavior in virtual execution environments. In: Proc. of the 12th Int'l Workshop on Recent Advances in Intrusion Detection. Saint-Malo: Springer, 2009. 304–325. [doi: 10.1007/978-3-642-04342-0_16]

[75] Steffens T. Attribution of Advanced Persistent Threats—How to Identify the Actors Behind Cyber-espionage. Berlin: Springer, 2020. [doi: 10.1007/978-3-662-61313-9]

[76] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proc. of the 26th Int'l Conf. on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2013. 3111–3119.

[77] Ding SHH, Fung BCM, Charland P. Asm2Vec: Boosting static representation robustness for binary clone search against code obfuscation and compiler optimization. In: Proc. of the 2019 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2019. 472–489.

[doi: 10.1109/SP.2019.00003]

[78] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations. In: Proc. of the 20th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM, 2014. 701–710. [doi: 10.1145/2623330.2623732]

[79] Lyu YQ, Wang ZY, Yang XZ, Song WN, Peng GJ. A novel APT malware classification method based on feature function code. Journal of Zhengzhou University (Natural Science Edition), 2023, 55(2): 10–17, 24. (in Chinese with English abstract). [doi: 10.13705/j.issn.1671-6841.2021417]

[80] Qian YC, Peng GJ, Wang Y, Liang Y. Homology analysis of malicious code and family clustering. Computer Engineering and Applications, 2015, 51(18): 76–81 (in Chinese with English abstract). [doi: 10.3778/j.issn.1002-8331.1411-0342]

[81] Antiy CERT. The dance of the White Elephant: Cyberattacks from the subcontinent. 2016 (in Chinese). https://www.antiy.cn/research/notice&report/research_report/304.html

[82] Kaspersky. Gauss: Abnormal distribution. 2012. https://securelist.com/gauss-abnormal-distribution/36620/

[83] Caliskan A, Yamaguchi F, Dauber E, Harang RE, Rieck K, Greenstadt R, Narayanan A. When coding style survives compilation: De-anonymizing programmers from executable binaries. In: Proc. of the 25th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2018. [doi: 10.14722/ndss.2018.23304]

[84] Alrabaee S, Karbab EB, Wang LY, Debbabi M. BinEye: Towards efficient binary authorship characterization using deep learning. In: Proc. of the 24th European Symp. on Research in Computer Security. Luxembourg: Springer, 2019. 47–67. [doi: 10.1007/978-3-030-29962-0_3]

[85] Han WJ, Xue JF, Wang Y, Zhang FQ, Gao XW. APTMalInsight: Identify and cognize APT malware based on system call information and ontology knowledge framework. Information Sciences, 2021, 546: 633–664. [doi: 10.1016/j.ins.2020.08.095]

[86] Ma SQ, Zhang XY, Xu DY. ProTracer: Towards practical provenance tracing by alternating between logging and tainting. In: Proc. of the 23rd Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2016. [doi: 10.14722/ndss.2016.23350]

[87] Rosenberg I, Sicard G, David E. DeepAPT: Nation-state APT attribution using end-to-end deep neural networks. In: Proc. of the 26th Int'l Conf. on Artificial Neural Networks. Alghero: Springer, 2017. 91–99. [doi: 10.1007/978-3-319-68612-7_11]

[88] Rosenberg I, Sicard G, David E. End-to-end deep neural networks and transfer learning for automatic analysis of nation-state malware. Entropy, 2018, 20(5): 390. [doi: 10.3390/e20050390]

[89] Gostev A, Kuznetsov I. Stuxnet/Duqu: The evolution of drivers. 2011. https://securelist.com/stuxnetduqu-the-evolution-of-drivers/36462

[90] Bencsáth B, Pék G, Buttyán L, Félegyházi M. The cousins of stuxnet: Duqu, flame, and gauss. Future Internet, 2012, 4(4): 971–1003. [doi: 10.3390/fi4040971]

[91] FireEye. APT28: A Window into Russia's cyber espionage operations? 2014. https://www.williamsnyder.net/wp-content/uploads/2015/04/FireEye-Russias-Cyber-Espionage-Report.pdf

[92] Li SD, Zhang QQ, Wu XB, Han WH, Tian ZH. Attribution classification method of APT malware in IoT using machine learning techniques. Security and Communication Networks, 2021, 2021: 9396141. [doi: 10.1155/2021/9396141]

[93] Zhao JJ. Research on author organization features of malware on Windows [MS. Thesis]. Guangzhou: Jinan University, 2020 (in Chinese with English abstract). [doi: 10.27167/d.cnki.gjinu.2020.000381]

[94] Laurenza G, Lazzeretti R, Mazzotti L. Malware triage for early identification of Advanced Persistent Threat activities. Digital Threats: Research and Practice, 2020, 1(3): 16. [doi: 10.1145/3386581]

[95] Wang QQ, Yan HB, Han ZH. Explainable APT attribution for malware using NLP techniques. In: Proc. of the 21st IEEE Int'l Conf. on Software Quality, Reliability and Security. Hainan: IEEE, 2021. 70–80. [doi: 10.1109/QRS54544.2021.00018]

[96] Checkpoint, Intezer Lab. Mapping the connections inside Russia's APT ecosystem. 2019. https://research.checkpoint.com/2019/Russianaptecosystem

[97] Li T, Qiao W, Zhang JW, Gao YY, Wang SN, Shen YL, Ma JF. Privacy-preserving network attack provenance based on graph convolutional neural network. Journal of Computer Research and Development, 2021, 58(5): 1006–1020 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2021.20200942]

[98] Krsul I, Spafford EH. Authorship analysis: Identifying the author of a program. Computers & Security, 1997, 16(3): 233–257. [doi: 10.1016/S0167-4048(97)00005-9]

[99] Burrows S, Uitdenbogerd A, Turpin A. Application of information retrieval techniques for source code authorship attribution. In: Proc. of the 14th Int'l Conf. on Database Systems for Advanced Applications. Brisban: Springer, 2009. 699–713. [doi: 10.1007/978-3-642-00887-0_61]

[100] Chen R, Hong LN, Lü C, Deng W. Author identification of software source code with program dependence graphs. In: Proc. of the 34th IEEE Annual Computer Software and Applications Conf. Workshops. Seoul: IEEE, 2010. 281–286. [doi: 10.1109/COMPSACW.2010.

56]

[101] Caliskan-Islam A, Harang R, Liu A, Narayanan A, Voss C, Yamaguchi F, Greenstadt R. De-anonymizing programmers via code stylometry. In: Proc. of the 24th USENIX Conf. on Security Symp. Washington: USENIX Association, 2015. 255–270.

[102] Dauber E, Caliskan A, Harang R, Shearer G, Weisman M, Nelson F, Greenstadt R. Git blame who? Stylistic authorship attribution of small, incomplete source code fragments. Proc. on Privacy Enhancing Technologies, 2019, 2019(3): 389–408. [doi: 10.2478/popets-2019-0053]

[103] Alsulami B, Dauber E, Harang R, Mancoridis S, Greenstadt R. Source code authorship attribution using long short-term memory based networks. In: Proc. of the 22nd European Symp. on Research in Computer Security. Oslo: Springer, 2017. 65–82. [doi: 10.1007/978-3-319-66402-6_6]

[104] Abuhamad M, AbuHmed T, Mohaisen A, Nyang D. Large-scale and language-oblivious code authorship identification. In: Proc. of the 2018 ACM SIGSAC Conf. on Computer and Communications Security. Toronto: ACM, 2018. 101–114. [doi: 10.1145/3243734.3243738]

[105] Rosenblum N, Zhu XJ, Miller BP. Who wrote this code? Identifying the authors of program binaries. In: Proc. of the 16th European Symp. on Research in Computer Security. Leuven: Springer, 2011. 172–189. [doi: 10.1007/978-3-642-23822-2_10]

[106] Marquis-Boire M, Marschalek M, Guarnieri C. Big game hunting: The peculiarities in nation-state malware research. In: Proc. of the 2015 Black Hat, 2015. https://www.blackhat.com/docs/us-15/materials/us-15-MarquisBoire-Big-Game-Hunting-The-Peculiarities-Of-Nation-State-malware-Research.pdf

[107] Alrabaee S, Saleem N, Preda S, Wang LY, Debbabi M. OBA2: An onion approach to binary code authorship attribution. Digital Investigation, 2014, 11(S1): S94–S103. [doi: 10.1016/j.diin.2014.03.012]

[108] Qiao YC, Yun XC, Zhang YZ, Li SH. An automatic malware homology identification method based on calling habits. Acta Electronica Sinica, 2016, 44(10): 2410–2414 (in Chinese with English abstract). [doi: 10.3969/j.issn.0372-2112.2016.10.019]

[109] Alrabaee S, Debbabi M, Wang LY. CPA: Accurate cross-platform binary authorship characterization using LDA. IEEE Trans. on Information Forensics and Security, 2020, 15: 3051–3066. [doi: 10.1109/TIFS.2020.2980190]

[110] Wu P. Research on homology determination technology of multi-form software code [Ph.D. Thesis]. Chengdu: Sichuan University, 2021 (in Chinese with English abstract). [doi: 10.27342/d.cnki.gscdu.2021.000007]

[111] Wang ZH. Research on traceability technology of ransomware [MS. Thesis]. Beijing: Beijing University of Posts and Telecommunications, 2019 (in Chinese with English abstract).

[112] Berady A, Jaume M, Tong VVT, Guette G. From TTP to IoC: Advanced persistent graphs for threat hunting. IEEE Trans. on Network and Service Management, 2021, 18(2): 1321–1333. [doi: 10.1109/TNSM.2021.3056999]

[113] Lu N, Zhang JW, Ma JF, Cong X, Shi WB, Wang SG. A scalable IP traceback approach employing dynamic deterministic packet marking in the large-scale networks. Chinese Journal of Computers, 2020, 43(8): 1493–1516 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2020.01493]

[114] Tian HC, Bi J. An incrementally deployable flow-based scheme for IP traceback. IEEE Communications Letters, 2012, 16(7): 1140–1143. [doi: 10.1109/LCOMM.2012.051512.120467]

[115] Xu K, Zhu L, Zhu M. Architecture and key technologies of Internet address security. Ruan Jian Xue Bao/Journal of Software, 2014, 25(1): 78–97 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4509.htm [doi: 10.13328/j.cnki.jos.004509]

[116] Yu S, Zhou WL, Guo S, Guo MY. A feasible IP traceback framework through dynamic deterministic packet marking. IEEE Trans. on Computers, 2016, 65(5): 1418–1427. [doi: 10.1109/TC.2015.2439287]

[117] Lu N, Zhang WJ, Ma JF, Cheng QF, Zhang JW, Wang SG. Efficient single-packet traceback approach based on alliance theory. Ruan Jian Xue Bao/Journal of Software, 2020, 31(12): 3880–3908 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5882.htm [doi: 10.13328/j.cnki.jos.005882]

[118] Jiang JG, Wang JZ, Kong B, Hu B, Liu JQ. On the survey of network attack source traceback. Journal of Cyber Security, 2018, 3(1): 111–131 (in Chinese with English abstract). [doi: 10.19363/j.cnki.cn10-1380/tn.2018.01.008]

[119] Hong KF, Chen CC, Chiu YT, Chou KS. Ctracer: Uncover C&C in advanced persistent threats based on scalable framework for enterprise log data. In: Proc. of the 2015 IEEE Int'l Congress on Big Data. New York: IEEE, 2015. 551–558. [doi: 10.1109/BigDataCongress.2015.86]

[120] Oprea A, Li Z, Yen TF, Chin SH, Alrwais S. Detection of early-stage enterprise infection by mining large-scale log data. In: Proc. of the 45th Annual IEEE/IFIP Int'l Conf. on Dependable Systems and Networks. Rio de Janeiro: IEEE, 2015. 45–56. [doi: 10.1109/DSN.2015.14]

[121] Fuller J, Kasturi RP, Sikder A, Xu HC, Arik B, Verma V, Asdar E, Saltaformaggio B. C3PO: Large-scale study of covert monitoring of C&C servers via over-permissioned protocol infiltration. In: Proc. of the 2021 ACM SIGSAC Conf. on Computer and Communications

Security. ACM, 2021. 3352–3365. [doi: 10.1145/3460120.3484537]
[122] Cohen I, Itkin E. Graphology of an Exploit-Hunting for exploits by looking for the author's fingerprints. 2020. https://research.checkpoint.com/2020/graphology-of-an-exploit-volodya/
[123] Irshad H, Ciocarlie G, Gehani A, Yegneswaran V, Lee KH, Patel J, Jha S, Kwon Y, Xu DY, Zhang XY. TRACE: Enterprise-wide provenance tracking for real-time APT detection. IEEE Trans. on Information Forensics and Security, 2021, 16: 4363–4376. [doi: 10.1109/TIFS.2021.3098977]
[124] Checkpoint. Huawei home routers in botnet recruitment. 2017. https://research.checkpoint.com/2017/good-zero-day-skiddie/
[125] Cook A, Nicholson A, Janicke H, Maglaras L, Smith R. Attribution of cyber attacks on industrial control systems. EAI Endorsed Trans. on Industrial Networks and Intelligent Systems, 2016, 3(7): e3. [doi: 10.4108/eai.21-4-2016.151158]
[126] Wheeler DA, Larsen GN. Techniques for cyber attack attribution. Technical Report: IDA Paper P-3792, Institute for Defense Analyses, 2003.
[127] Antiy. Antiy capture system. 2022 (in Chinese). https://www.antiy.cn/Security_Product/ACS.html
[128] Qianxin. Attacks honeypot and hunting system. 2022 (in Chinese). https://www.qianxin.com/product/detail/pid/426
[129] Trellix. The trellix XDR platform. 2022. https://www.trellix.com/en-us/products.html
[130] Dai JR, Zhang Y, Xu HL, Lyu HM, Wu ZC, Xing XY, Yang M. Facilitating vulnerability assessment through PoC migration. In: Proc. of the 2021 ACM SIGSAC Conf. on Computer and Communications Security. ACM, 2021. 3300–3317. [doi: 10.1145/3460120.3484594]
[131] Cohen D, Narayanaswamy K. Attack attribution in non-cooperative networks. In: Proc. of the 5th Annual IEEE SMC Information Assurance Workshop. West Point: IEEE, 2004. 436–437. [doi: 10.1109/IAW.2004.1437851]
[132] Qianxin. Level analysis of threat intelligence. 2016 (in Chinese). https://ti.qianxin.com/blog/articles/level-of-threat-intelligence/
[133] Leng T, Cai LJ, Yu AM, Zhu ZY, Ma J, Li CF, Niu RC, Meng D. Review of threat discovery and forensic analysis based on system provenance graph. Journal on Communications, 2022, 43(7): 172–188 (in Chinese with English abstract). [doi: 10.11959/j.issn.1000-436x.2022105]
[134] Hassan WU, Bates A, Marino D. Tactical provenance analysis for endpoint detection and response systems. In: Proc. of the 2020 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2020. 1172–1189. [doi: 10.1109/SP40000.2020.00096]
[135] Afianian A, Niksefat S, Sadeghiyan B, Baptiste D. malware dynamic analysis evasion techniques: A survey. ACM Computing Surveys, 2020, 52(6): 126. [doi: 10.1145/3365001]
[136] Zhang J, Peng GJ, Yang XZ. Malicious evasion sample detection based on dynamic API call sequence and machine learning. Journal of Shandong University (Natural Science), 2022, 57(7): 85–93, 102 (in Chinese with English abstract). [doi: 10.6040/j.issn.1671-9352.2.2021.117]
[137] Kirat D, Vigna G, Kruegel C. BareCloud: Bare-metal analysis-based evasive malware detection. In: Proc. of the 23rd USENIX Security Symp. San Diego: USENIX Association, 2014. 287–301.
[138] Zynamics. BinDiff. 2016. https://www.zynamics.com/bindiff.html
[139] Duan Y, Li XZX, Wang JH, Yin H. DeepBinDiff: Learning program-wide code representations for binary diffing. In: Proc. of the 27th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2020. [doi: 10.14722/ndss.2020.24311]
[140] Kumar T, Somani G. Origin information assisted hybrid analysis to detect APT malware. In: Proc. of the 17th Int'l Conf. on Information Systems Security. Patna: Springer, 2021. 75–93. [doi: 10.1007/978-3-030-92571-0_5]
[141] Nataraj L, Karthikeyan S, Jacob G, Manjunath BS. malware images: Visualization and automatic classification. In: Proc. of the 8th Int'l Symp. on Visualization for Cyber Security. Pittsburgh: ACM, 2011. 4. [doi: 10.1145/2016904.2016908]
[142] Do Xuan C, Dao MH. A novel approach for APT attack detection based on combined deep learning model. Neural Computing and Applications, 2021, 33(20): 13251–13264. [doi: 10.1007/s00521-021-05952-5]
[143] Shang LK, Guo D, Ji YD, Li Q. Discovering unknown advanced persistent threat using shared features mined by neural networks. Computer Networks, 2021, 189: 107937. [doi: 10.1016/j.comnet.2021.107937]
[144] Wei CX, Li Q, Guo D, Meng XY. Toward identifying APT malware through API system calls. Security and Communication Networks, 2021: 8077220. [doi: 10.1155/2021/8077220]
[145] Gori M, Monfardini G, Scarselli F. A new model for learning in graph domains. In: Proc. of the 2005 IEEE Int'l Joint Conf. on Neural Networks. Montreal: IEEE, 2005. 729–734. [doi: 10.1109/IJCNN.2005.1555942]
[146] Wang YQ. Principles and Methods of Artificial Intelligence. Xi'an: Xi'an Jiaotong University Press, 1998 (in Chinese).
[147] Tari L. Knowledge inference. In: Dubitzky W, Wolkenhauer O, Cho KH, Yokota H, eds. Encyclopedia of Systems Biology. New York: Springer, 2013. 1074–1078. [doi: 10.1007/978-1-4419-9863-7_166]
[148] Ren HY, Dai HJ, Dai B, Chen XY, Yasunaga M, Sun HT, Schuurmans D, Leskovec J, Zhou D. LEGO: Latent execution-guided

reasoning for multi-hop question answering on knowledge graphs. In: Proc. of the 38th Int'l Conf. on Machine Learning. PMLR, 2021. 8959–8970.

[149] Chen ZH, Li GB, Wan X. Align, reason and learn: Enhancing medical vision-and-language pre-training with knowledge. In: Proc. of the 30th ACM Int'l Conf. on Multimedia. Lisboa: ACM, 2022. 5152–5161. [doi: 10.1145/3503161.3547948]

[150] Marin E, Almukaynizi M, Shakarian P. Reasoning about future cyber-attacks through socio-technical hacking information. In: Proc. of the 31st IEEE Int'l Conf. on Tools with Artificial Intelligence. Portland: IEEE, 2019. 157–164. [doi: 10.1109/ICTAI.2019.00030]

[151] Zhang SQ, Bai GY, Li H, Zhang MZ. IoT security knowledge reasoning method of multi-source data fusion. Journal of Computer Research and Development, 2022, 59(12): 2735–2749 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.20210954]

[152] Chong S, Van Der Meyden R. Using architecture to reason about information security. ACM Trans. on Information and System Security, 2015, 18(2): 8. [doi: 10.1145/2829949]

[153] Paja E, Dalpiaz F, Giorgini P. Modelling and reasoning about security requirements in socio-technical systems. Data & Knowledge Engineering, 2015, 98: 123–143. [doi: 10.1016/j.datak.2015.07.007]

[154] Peralta M, Mukhopadhyay S, Bharadwaj R. Reasoning about security in sensor networks. Concurrency and Computation: Practice and Experience, 2015, 27(15): 3816–3841. [doi: 10.1002/cpe.3433]

[155] Channakeshava K, Bisset K, Marathe MV, Kumar A, Vullikanti S. Reasoning about mobile malware using high performance computing based population scale models. In: Proc. of the 2014 Winter Simulation Conf. Savannah: IEEE, 2014. 3048–3059. [doi: 10.1109/WSC.2014.7020143]

[156] Marin E, Almukaynizi M, Shakarian P. Inductive and deductive reasoning to assist in cyber-attack prediction. In: Proc. of the 10th Annual Computing and Communication Workshop and Conf. Las Vegas: IEEE, 2020. 262–268. [doi: 10.1109/CCWC47524.2020.9031154]

[157] Xing QQ. Network security risk assessment based on intelligent planning [MS. Thesis]. Changsha: National University of Defense Technology, 2014 (in Chinese with English abstract).

[158] Karafili E, Wang LN, Lupu EC. An argumentation-based reasoner to assist digital investigation and attribution of cyber-attacks. Forensic Science International: Digital Investigation, 2020, 32: 300925. [doi: 10.1016/j.fsidi.2020.300925]

[159] Ghosh K, Morales JA, Casey W, Mishra B. Sandboxing and reasoning on malware infection trees. In: Proc. of the 10th Int'l Conf. on Malicious and Unwanted Software. Fajardo: IEEE, 2015. 69–73. [doi: 10.1109/MALWARE.2015.7413686]

[160] Klobas JE, McGill T, Wang XQ. How perceived security risk affects intention to use smart home devices: A reasoned action explanation. Computers & Security, 2019, 87: 101571. [doi: 10.1016/j.cose.2019.101571]

[161] Cauli C, Li M, Piterman N, Tkachuk O. Pre-deployment security assessment for cloud services through semantic reasoning. In: Proc. of the 33rd Int'l Conf. on Computer Aided Verification. Springer, 2021. 767–780. [doi: 10.1007/978-3-030-81685-8_36]

[162] Ochoa M, Banescu S, Disenfeld C, Barthe G, Ganesh V. Reasoning about probabilistic defense mechanisms against remote attacks. In: Proc. of the 2017 IEEE European Symp. on Security and Privacy. Paris: IEEE, 2017. 499–513. [doi: 10.1109/EuroSP.2017.30]

[163] Ma PC, Jiang B, Lu ZG, Li N, Jiang ZW. Cybersecurity named entity recognition using bidirectional long short-term memory with conditional random fields. Tsinghua Science and Technology, 2021, 26(3): 259–265. [doi: 10.26599/TST.2019.9010033]

[164] Yang XZ, Peng GJ, Li ZC, Lyu Y, Liu SD, Li CG. Research on entity recognition and alignment of APT attack based on BERT and BiLSTM-CRF. Journal on Communications, 2022, 43(6): 58–70 (in Chinese with English abstract). [doi: 10.11959/j.issn.1000−436x.2022116]

[165] Solic K, Ocevcic H, Golub M. The information systems' security level assessment model based on an ontology and evidential reasoning approach. Computers & Security, 2015, 55: 100–112. [doi: 10.1016/j.cose.2015.08.004]

[166] Wu SY, Zhang Y, Cao W. Network security assessment using a semantic reasoning and graph based approach. Computers & Electrical Engineering, 2017, 64: 96–109. [doi: 10.1016/j.compeleceng.2017.02.001]

[167] Rastogi N, Dutta S, Gittens A, Zaki MJ, Aggarwal C. TINKER: A framework for open source cyberthreat intelligence. In: Proc. of the 2022 IEEE Int'l Conf. on Trust, Security and Privacy in Computing and Communications. Wuhan: IEEE, 2022. 1569–1574. [doi: 10.1109/TrustCom56396.2022.00225]

[168] Zhang H, Yao DF, Ramakrishnan N, Zhang ZB. Causality reasoning about network events for detecting stealthy malware activities. Computers & Security, 2016, 58: 180–198. [doi: 10.1016/j.cose.2016.01.002]

[169] He ZC, Hu GY, Lee R. New models for understanding and reasoning about speculative execution attacks. In: Proc. of the 2021 IEEE Int'l Symp. on High-performance Computer Architecture. Seoul: IEEE, 2021. 40–53. [doi: 10.1109/HPCA51647.2021.00014]

[170] Alserhani FM. Knowledge-based model to represent security information and reason about multi-stage attacks. In: Proc. of the 2015 Int'l Conf. on Advanced Information Systems Engineering. Stockholm: Springer, 2015. 482–494. [doi: 10.1007/978-3-319-19243-7_44]

[171] Lansley M, Polatidis N, Kapetanakis S, Amin K, Samakovitis G, Petridis M. Seen the villains: Detecting social engineering attacks

using case-based reasoning and deep learning. In: Proc. of the 27th Int'l Conf. on Case-based Reasoning Workshop. Otzenhausen: CEUR-WS.org, 2019. 39–48.

[172] Zhao J, Yan QB, Liu XD, Li B, Zuo GS. Cyber threat intelligence modeling based on heterogeneous graph convolutional network. In: Proc. of the 23rd Int'l Symp. on Research in Attacks, Intrusions and Defenses. San Sebastian: USENIX Association, 2020. 241–256.

[173] Zhang SQ, Bai GY, Li H, Liu PP, Zhang MZ, Li SJ. Multi-source knowledge reasoning for data-driven IoT security. Sensors, 2021, 21(22): 7579. [doi: 10.3390/s21227579]

[174] Zengy J, Wang X, Liu JH, Chen YF, Liang ZK, Chua TS, Chua ZL. Shadewatcher: Recommendation-guided cyber threat analysis using system audit records. In: Proc. of the 2022 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2022. 489–506. [doi: 10.1109/SP46214.2022.9833669]

[175] Skowyra R, Gomez SR, Bigelow D, Landry J, Okhravi H. QUASAR: Quantitative attack space analysis and reasoning. In: Proc. of the 33rd Annual Computer Security Applications Conf. Orlando: ACM, 2017. 68–78. [doi: 10.1145/3134600.3134633]

[176] Alsaheel A, Nan YH, Ma SQ, Yu L, Walkup G, Celik ZB, Zhang XY, Xu DY. ATLAS: A sequence-based learning approach for attack investigation. In: Proc. of the 30th USENIX Security Symp. USENIX Association, 2021. 3005–3022.

[177] Sigholm J, Bang M. Towards offensive cyber counterintelligence: Adopting a target-centric view on advanced persistent threats. In: Proc. of the 2013 European Intelligence and Security Informatics Conf. Uppsala: IEEE, 2013. 166–171. [doi: 10.1109/EISIC.2013.37]

[178] Jia SS, Xu YB. The APT detection method based on attack tree for SDN. In: Proc. of the 2nd Int'l Conf. on Cryptography, Security and Privacy. Guiyang: ACM, 2018. 116–121. [doi: 10.1145/3199478.3199481]

[179] Akbar KA, Wang YG, Islam S, Singhal A, Khan L, Thuraisingham B. Identifying tactics of advanced persistent threats with limited attack traces. In: Proc. of the 17th Int' Conf. on Information Systems Security. Patna: Springer, 2021. 3–25. [doi: 10.1007/978-3-030-92571-0_1]

[180] Chen WX. A research on attack-path prediction method for APT organization [MS. Thesis]. Guangzhou: Guangzhou University, 2021 (in Chinese with English abstract). [doi: 10.27040/d.cnki.ggzdu.2021.000867]

[181] Zhang YF, Zhang ZY, Qu HK, Zhang G, Wang ZB, Wang BL. Key path analysis method for large-scale industrial control network. Chinese Journal of Network and Information Security, 2021, 7(6): 31–43 (in Chinese with English abstract). [doi: 10.11959/j.issn.2096-109x.2021069]

[182] Li SY, Cui BJ. Research on association analysis technology of network attack trace based on Web log. In: Barolli L, Poniszewska-Maranda A, Park H, eds. Innovative Mobile and Internet Services in Ubiquitous Computing. Cham: Springer, 2021. 33–43. [doi: 10.1007/978-3-030-50399-4_4]

[183] Stellios I, Kotzanikolaou P, Grigoriadis C. Assessing IoT enabled cyber-physical attack paths against critical systems. Computers & Security, 2021, 107: 102316. [doi: 10.1016/j.cose.2021.102316]

[184] Phillips C, Swiler LP. A graph-based system for network-vulnerability analysis. In: Proc. of the 1998 workshop on New Security Paradigms. Charlottesville: ACM, 1998. 71–79. [doi: 10.1145/310889.310919]

[185] Hu H, Liu YL, Zhang HQ, Yang YJ, Ye RG. Route prediction method for network intrusion using absorbing Markov chain. Journal of Computer Research and Development, 2018, 55(4): 831–845 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2018.20170087]

[186] Yu Y, Xia CH, Hu XY. Defense scheme generation method using mixed path attack graph. Journal of Zhejiang University (Engineering Science), 2017, 51(9): 1745–1759 (in Chinese with English abstract). [doi: 10.3785/j.issn.1008-973X.2017.09.009]

[187] Sadlek L, Čeleda P, Tovarňák D. Identification of attack paths using kill chain and attack graphs. In: Proc. of the 2022 IEEE/IFIP Network Operations and Management Symp. Budapest: IEEE, 2022. 1–6. [doi: 10.1109/NOMS54207.2022.9789803]

[188] Zhai HX, Lu YM, Wang H, Ao S. Research on attack path prediction method based on T_NAG model. Application Research of Computers, 2021, 38(3): 886–892 (in Chinese with English abstract). [doi: 10.19734/j.issn.1001-3695.2019.11.0701]

[189] Sun C, Hu H, Yang YJ, Zhang HQ. Prediction method of 0day attack path based on cyber defense knowledge graph. Chinese Journal of Network and Information Security, 2022, 8(1): 151–166 (in Chinese with English abstract). [doi: 10.11959/j.issn.2096-109x.2021101]

[190] Dai J, Sun XY, Liu P. Patrol: Revealing zero-day attack paths through network-wide system object dependencies. In: Proc. of the 18th European Symp. on Research in Computer Security. Egham: Springer, 2013. 536–555. [doi: 10.1007/978-3-642-40203-6_30]

[191] Lu JZ. APT attack modeling and detection technology based on depth analysis of whole network traffic and logs [Ph.D. Thesis]. Chengdu: University of Electronic Science and Technology of China, 2019 (in Chinese with English abstract).

[192] Yang XY, Liu HY, Wang ZY, Gao P. Zebra: Deeply integrating system-level provenance search and tracking for efficient attack investigation. arXiv:2211.05403, 2022.

[193] Kim T, Park N, Hong J, Kim SW. Phishing URL detection: A network-based approach robust to evasion. In: Proc. of the 2022 ACM SIGSAC Conf. on Computer and Communications Security. Los Angeles: ACM, 2022. 1769–1782. [doi: 10.1145/3548606.3560615]

[194] Shu XK, Yao DF, Ramakrishnan N. Unearthing stealthy program attacks buried in extremely long execution paths. In: Proc. of the 22nd ACM SIGSAC Conf. on Computer and Communications Security. Denver: ACM, 2015. 401–413. [doi: 10.1145/2810103.2813654]

[195] Sun XY, Dai J, Liu P, Singhal A, Yen J. Using Bayesian networks for probabilistic identification of zero-day attack paths. IEEE Trans. on Information Forensics and Security, 2018, 13(10): 2506–2521. [doi: 10.1109/TIFS.2018.2821095]

[196] Luo HB, Chen Z, Li JW, Vasilakos AV. Preventing distributed denial-of-service flooding attacks with dynamic path identifiers. IEEE Trans. on Information Forensics and Security, 2017, 12(8): 1801–1815. [doi: 10.1109/TIFS.2017.2688414]

[197] Gursoy ME, Liu L, Truex S, Yu L, Wei WQ. Utility-aware synthesis of differentially private and attack-resilient location traces. In: Proc. of the 2018 ACM SIGSAC Conf. on Computer and Communications Security. Toronto: ACM, 2018. 196–211. [doi: 10.1145/3243734.3243741]

[198] Li FH, Li YJ, Leng SY, Guo YC, Geng K, Wang Z, Fang L. Dynamic countermeasures selection for multi-path attacks. Computers & Security, 2020, 97: 101927. [doi: 10.1016/j.cose.2020.101927]

[199] Ma LR, Yang L, Wang JX, Tang X. Using fuzzy clustering to reconstruct alert correlation graph of intrusion detection. Journal on Communications, 2006, 27(9): 47–52 (in Chinese with English abstract). [doi: 10.3321/j.issn:1000-436X.2006.09.008]

[200] Fu X, Shi J, Xie L. Layered intrusion scenario reconstruction method for automated evidence analysis. Ruan Jian Xue Bao/Journal of Software, 2011, 22(5): 996–1008 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/3759.htm [doi: 10.3724/SP.J.1001.2011.03759]

[201] Tajima Y, Koide H. Applying the attacks tracer on advanced persistent threats to real networks. In: Proc. of the 9th Int'l Symp. on Computing and Networking Workshops. Matsue: IEEE, 2021. 392–397. [doi: 10.1109/CANDARW53999.2021.00072]

[202] Dain O, Cunningham RK. Fusing a heterogeneous alert stream into scenarios. In: Barbará D, Jajodia S, eds. Applications of Data Mining in Computer Security. Boston: Springer, 2002. 103–122. [doi: 10.1007/978-1-4615-0953-0_5]

[203] Wang S, Tang GM, Wang JH, Sun YF, Kou G. Attack scenario construction method based on causal knowledge net. Journal of Computer Research and Development, 2018, 55(12): 2620–2636 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2018.20160940]

[204] Hossain N, Sheikhi S, Sekar R. Combating dependence explosion in forensic analysis using alternative tag propagation semantics. In: Proc. of the 2020 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2020. 1139–1155. [doi: 10.1109/SP40000.2020.00064]

[205] Hassan WU, Guo SJ, Li D, Chen ZZ, Jee K, Li ZC, Bates A. NoDoze: Combatting threat alert fatigue with automated provenance triage. In: Proc. of the 26th Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2019. 1–15. [doi: 10.14722/ndss.2019.23349]

[206] Fang PC, Gao P, Liu CL, Ayday E, Jee K, Wang T, Ye YF, Liu ZT, Xiao XS. Back-propagating system dependency impact for attack investigation. In: Proc. of the 31st USENIX Security Symp. Boston: USENIX Association, 2022. 2461–2478.

[207] Ju AK. Research on key technologies of targeted cyber attacks detection based on multi-source heterogeneous data [Ph.D. Thesis]. Zhengzhou: Information Engineering University, 2020 (in Chinese with English abstract).

[208] Zhu ZY, Dumitras T. ChainSmith: Automatically learning the semantics of malicious campaigns by mining threat intelligence reports. In: Proc. of the 2018 IEEE European Symp. on Security and Privacy. London: IEEE, 2018. 458–472. [doi: 10.1109/EuroSP.2018.00039]

[209] Satvat K, Gjomemo R, Venkatakrishnan VN. Extractor: Extracting attack behavior from threat reports. In: Proc. of the 2021 IEEE European Symp. on Security and Privacy. Vienna: IEEE, 2021. 598–615. [doi: 10.1109/EuroSP51992.2021.00046]

[210] Shen Y, Stringhini G. ATTACK2VEC: Leveraging temporal word embeddings to understand the evolution of cyberattacks. In: Proc. of the 28th USENIX Conf. on Security Symp. Santa Clara: USENIX Association, 2019. 905–921.

[211] Tam K, Khan SJ, Fattori A, Cavallaro L. CopperDroid: Automatic reconstruction of Android malware behaviors. In: Proc. of the 22nd Annual Network and Distributed System Security Symp. San Diego: The Internet Society, 2015. [doi: 10.14722/ndss.2015.23145]

[212] Gao XE, Chen B, Jiang PL, Xiang ZT, Chen YF, Wang YM. Hierarchy-entropy based method for command and control networks reconfiguration. The Journal of Supercomputing, 2022, 78(13): 15229–15249. [doi: 10.1007/s11227-022-04445-z]

[213] Salem A, Bhattacharya A, Backes M, Fritz M, Zhang Y. Updates-leak: Data set inference and reconstruction attacks in online learning. In: Proc. of the 29th USENIX Conf. on Security Symp. USENIX Association, 2020. 73.

[214] Mao BF, Liu J, Lai YX, Sun MT. MIF: A multi-step attack scenario reconstruction and attack chains extraction method based on multi-information fusion. Computer Networks, 2021, 198: 108340. [doi: 10.1016/j.comnet.2021.108340]

[215] Gartner. Consider deception as a defense strategy against attackers. 2016. https://www.gartner.com/smarterwithgartner/deception-wave

[216] Jia ZP, Fang BX, Liu CG, Liu QX, Lin JB. Survey on cyber deception. Journal on Communications, 2017, 38(12): 128–143 (in Chinese with English abstract). [doi: 10.11959/j.issn.1000-436x.2017281]

[217] Cohen F. A note on the role of deception in information protection. Computers & Security, 1998, 17(6): 483–506. [doi: 10.1016/S0167-4048(98)80071-0]

[218] Araujo F, Hamlen KW, Biedermann S, Katzenbeisser S. From patches to honey-patches: Lightweight attacker misdirection, deception, and disinformation. In: Proc. of the 2014 ACM SIGSAC Conf. on Computer and Communications Security. Scottsdale: ACM, 2014. 942–953. [doi: 10.1145/2660267.2660329]

[219] Yang JL, Zhou Y, Zhou JS. Development of intelligent deception and anti-deception technology in the U.S. army. National Defense Technology, 2022, 43(4): 48–53 (in Chinese with English abstract). [doi: 10.13943/j.issn1671-4547.2022.04.09]

[220] Zhuge JW, Tang Y, Han XH, Duan HX. Honeypot technology research and application. Ruan Jian Xue Bao/Journal of Software, 2013, 24(4): 825–842 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4369.htm [doi: 10.3724/SP.J.1001.2013.04369]

[221] The Honeynet Project. Know Your Enemy: Learning About Security Threats. 2nd ed. Boston: Addison-Wesley Professional, 2004.

[222] Spitzner L. Honeypot farms. 2003. http://www.symantec.com/connect/articles/honeypot-farms

[223] Zhan ZX, Xu MC, Xu SH. Characterizing honeypot-captured cyber attacks: Statistical framework and case study. IEEE Trans. on Information Forensics and Security, 2013, 8(11): 1775–1789. [doi: 10.1109/TIFS.2013.2279800]

[224] Akiyama M, Yagi T, Yada T, Mori T, Kadobayashi Y. Analyzing the ecosystem of malicious URL redirection through longitudinal observation from honeypots. Computers & Security, 2017, 69: 155–173. [doi: 10.1016/j.cose.2017.01.003]

[225] Irvene C, Formby D, Litchfield S, Beyah R. HoneyBot: A honeypot for robotic systems. Proc. of the IEEE, 2018, 106(1): 61–70. [doi: 10.1109/JPROC.2017.2748421]

[226] Fan WJ, Du ZH, Smith-Creasey M, Fernandez D. HoneyDOC: An efficient honeypot architecture enabling all-round design. IEEE Journal on Selected Areas in Communications, 2019, 37(3): 683–697. [doi: 10.1109/JSAC.2019.2894307]

[227] Horák K, Bošanský B, Tomášek P, Kiekintveld C, Kamhoua C. Optimizing honeypot strategies against dynamic lateral movement using partially observable stochastic games. Computers & Security, 2019, 87: 101579. [doi: 10.1016/j.cose.2019.101579]

[228] Tian W, Du M, Ji XP, Liu GJ, Dai YW, Han Z. Honeypot detection strategy against advanced persistent threats in industrial Internet of Things: A prospect theoretic game. IEEE Internet of Things Journal, 2021, 8(24): 17372–17381. [doi: 10.1109/JIOT.2021.3080527]

[229] Jafarian JH, Al-Shaer E, Duan Q. Adversary-aware IP address randomization for proactive agility against sophisticated attackers. In: Proc. of the 2015 IEEE Conf. on Computer Communications. Hong Kong: IEEE, 2015. 738–746. [doi: 10.1109/INFOCOM.2015.7218443]

[230] Dai B, Zhang ZH, Wang L, Liu Y. APT attack heuristic induction honeypot platform based on Snort and OpenFlow. In: Proc. of the 6th Smart Computing and Communication. New York City: Springer, 2022. 340–351. [doi: 10.1007/978-3-030-97774-0_31]

[231] Veluchamy S, Kathavarayan RS. Deep reinforcement learning for building honeypots against runtime DoS attack. Int'l Journal of Intelligent Systems, 2022, 37(7): 3981–4007. [doi: 10.1002/int.22708]

[232] Chen JY, Hu SL, Xing CY, Zhang GM. Deception defense method against intelligent penetration attack. Journal on Communications, 2022, 43(10): 106–120 (in Chinese with English abstract). [doi: 10.11959/j.issn.1000−436x.2022202]

[233] Yang LX, Li PD, Zhang YS, Yang XF, Xiang Y, Zhou WL. Effective repair strategy against advanced persistent threat: A differential game approach. IEEE Trans. on Information Forensics and Security, 2019, 14(7): 1713–1728. [doi: 10.1109/TIFS.2018.2885251]

[234] Song YB, Fan M, Yang JJ, Hu AQ. Multipath solution and blocking method of network attack traffic based on topology analysis. Netinfo Security, 2020, 20(3): 9–17 (in Chinese with English abstract). [doi: 10.3969/j.issn.1671-1122.2020.03.002]

[235] López-Morales E, Rubio-Medrano C, Doupé A, Shoshitaishvili Y, Wang RY, Bao T, Ahn GJ. HoneyPLC: A next-generation honeypot for industrial control systems. In: Proc. of the 2020 ACM SIGSAC Conf. on Computer and Communications Security. ACM, 2020. 279–291. [doi: 10.1145/3372297.3423356]

[236] Hajimaghsoodi M, Jalili R. RAD: A statistical mechanism based on behavioral analysis for DDoS attack countermeasure. IEEE Trans. on Information Forensics and Security, 2022, 17: 2732–2745. [doi: 10.1109/TIFS.2022.3172598]

[237] Qu DP, Lv GX, Qu SJ, Shen HY, Yang Y, Heng ZY. An effective and lightweight countermeasure scheme to multiple network attacks in NDN. IEEE/ACM Trans. on Networking, 2022, 30(2): 515–528. [doi: 10.1109/TNET.2021.3121001]

[238] Ahmad A, Webb J, Desouza KC, Boorman J. Strategically-motivated advanced persistent threat: Definition, process, tactics and a disinformation model of counterattack. Computers & Security, 2019, 86: 402–418. [doi: 10.1016/j.cose.2019.07.001]

[239] Singh S, Sharma PK, Moon SY, Moon D, Park JH. A comprehensive study on APT attacks and countermeasures for future networks and communications: Challenges and solutions. The Journal of Supercomputing, 2019, 75(8): 4543–4574. [doi: 10.1007/s11227-016-1850-4]

[240] Cheng BL, Ming J, Fu JM, Peng GJ, Chen T, Zhang XS, Marion JY. Towards paving the way for large-scale windows malware analysis: Generic binary unpacking with orders-of-magnitude performance boost. In: Proc. of the 2018 ACM SIGSAC Conf. on Computer and Communications Security. Toronto: ACM, 2018. 395–411. [doi: 10.1145/3243734.3243771]

[241] Sharma A, Gupta BB, Singh AK, Saraswat VK. Orchestration of APT malware evasive manoeuvers employed for eluding anti-virus and sandbox defense. Computers & Security, 2022, 115: 102627. [doi: 10.1016/j.cose.2022.102627]

[242] Nappa A, Johnson R, Bilge L, Caballero J, Dumitras T. The attack of the clones: A study of the impact of shared code on vulnerability

patching. In: Proc. of the 2015 IEEE Symp. on Security and Privacy. San Jose: IEEE, 2015. 692–708. [doi: 10.1109/SP.2015.48]

[243] Chen X, Andersen J, Mao ZM, Bailey M, Nazario J. Towards an understanding of anti-virtualization and anti-debugging behavior in modern malware. In: Proc. of the 2008 IEEE Int'l Conf. on Dependable Systems and Networks. Anchorage: IEEE, 2008. 177–186. [doi: 10.1109/DSN.2008.4630086]

[244] Galloro N, Polino M, Carminati M, Continella A, Zanero S. A systematical and longitudinal study of evasive behaviors in Windows malware. Computers & Security, 2022, 113: 102550. [doi: 10.1016/j.cose.2021.102550]

[245] Liu SS, Feng PB, Wang S, Sun K, Cao JH. Enhancing malware analysis sandboxes with emulated user behavior. Computers & Security, 2022, 115: 102613. [doi: 10.1016/j.cose.2022.102613]

[246] Ji SL, Du TY, Li JF, Shen C, Li B. Security and privacy of machine learning models: A survey. Ruan Jian Xue Bao/Journal of Software, 2021, 32(1): 41–67 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6131.htm [doi: 10.13328/j.cnki.jos.006131]

[247] Wang BL, Yao YS, Shan S, Li HY, Viswanath B, Zheng HT, Zhao BY. Neural cleanse: Identifying and mitigating backdoor attacks in neural networks. In: Proc. of the 2019 IEEE Symp. on Security and Privacy. San Francisco: IEEE, 2019. 707–723. [doi: 10.1109/SP.2019.00031]

[248] United States Cyber Command. Achieve and maintain cyberspace superiority. 2018. https://www.cybercom.mil/Portals/56/Documents/USCYBERCOMVisionApril2018.pdf

[249] Smeets M. U.S. cyber strategy of persistent engagement & defend forward: Implications for the alliance and intelligence collection. Intelligence and National Security, 2020, 35(3): 444–453. [doi: 10.1080/02684527.2020.1729316]

[250] Yang XZ, Peng GJ, Zhang DN, Gao YH, Li CG. PowerDetector: Malicious powershell script family classification based on multi-modal semantic fusion and deep learning. China Communications, 2023, 20(11): 202–224. [doi: 10.23919/JCC.fa.2022-0509.202311]

附中文参考文献:

[3] 付钰, 李洪成, 吴晓平, 王甲生. 基于大数据分析的 APT 攻击检测研究综述. 通信学报, 2015, 36(11): 1–14. [doi: 10.11959/j.issn.1000-436x.2015184]

[5] 陈瑞东, 张小松, 牛伟纳, 蓝皓月. APT 攻击检测与反制技术体系的研究. 电子科技大学学报, 2019, 48(6): 870–879. [doi: 10.3969/j.issn.1001-0548.2019.06.011]

[7] 安天实验室. 乌克兰电力系统遭受攻击事件综合分析报告. 2016. https://www.antiy.com/response/A_Comprehensive_Analysis_Report_on_Ukraine_Power_Grid_Outage/A_Comprehensive_Analysis_Report_on_Ukraine_Power_Grid_Outage.html

[9] 360 数字安全. 关于西北工业大学发现美国 NSA 网络攻击调查报告 (之一). 2022. https://mp.weixin.qq.com/s/0ReOzQMM5GS4xXRUPpKCvA

[11] 张瑜, 潘小明, Liu QZ, 曹均阔, 罗自强. APT 攻击与防御. 清华大学学报 (自然科学版), 2017, 57(11): 1127–1133. [doi: 10.16511/j.cnki.qhdxxb.2017.21.024]

[13] 王蕊, 冯登国, 杨轶, 苏璞睿. 基于语义的恶意代码行为特征提取及检测方法. 软件学报, 2012, 23(2): 378–393. http://www.jos.org.cn/1000-9825/3953.htm [doi: 10.3724/SP.J.1001.2012.03953]

[14] 宋文纳, 彭国军, 傅建明, 张焕国, 陈施旅. 恶意代码演化与溯源技术研究. 软件学报, 2019, 30(8): 2229–2267. http://www.jos.org.cn/1000-9825/5767.htm [doi: 10.13328/j.cnki.jos.005767]

[15] 杨秀璋, 彭国军, 罗元, 宋文纳, 张杰, 操方涛. OMRDetector: 一种基于深度学习的混淆恶意请求检测方法. 计算机学报, 2022, 45(10): 2167–2189. [doi: 10.11897/SP.J.1016.2022.02167]

[21] 官赛萍, 靳小龙, 贾岩涛, 王元卓, 程学旗. 面向知识图谱的知识推理研究进展. 软件学报, 2018, 29(10): 2966–2994. http://www.jos.org.cn/1000-9825/5551.htm [doi: 10.13328/j.cnki.jos.005551]

[23] 马钰锡, 张全新, 谭毓安, 沈蒙. 面向智能攻击的行为预测研究. 软件学报, 2021, 32(5): 1526–1546. http://www.jos.org.cn/1000-9825/6204.htm [doi: 10.13328/j.cnki.jos.006204]

[32] 李昂. 基于恶意代码基因的攻击组织特征提取方法研究 [硕士学位论文]. 北京: 北京邮电大学, 2021. [doi: 10.26969/d.cnki.gbydu.2021.001346]

[33] 黄克振, 连一峰, 冯登国, 张海霞, 吴迪, 马向亮. 一种基于图模型的网络攻击溯源方法. 软件学报, 2022, 33(2): 683–698. http://www.jos.org.cn/1000-9825/6314.htm [doi: 10.13328/j.cnki.jos.006314]

[39] 刘潮歌, 方滨兴, 刘宝旭, 崔翔, 刘奇旭. 定向网络攻击追踪溯源层次化模型研究. 信息安全学报, 2019, 4(4): 1–18. [doi: 10.19363/J.cnki.cn10-1380/tn.2019.07.01]

[40] 潘亚峰, 朱俊虎, 周天阳. APT 攻击场景重构方法综述. 信息工程大学学报, 2021, 22(1): 55–60, 80. [doi: 10.3969/j.issn.1671-0673.2021.01.010]

[42] 360 网络安全响应中心. 2021 年上半年全球高级持续性威胁 (APT) 研究报告. 2021. https://cert.360.cn/report/detail?id=6c9a1b56e4ceb84a8ab9e96044429adc

[48] 360 烽火实验室. Darkhotel(APT-C-06) 使用“双星”0Day 漏洞 (CVE-2019-17026、CVE-2020-0674) 针对中国发起的 APT 攻击分析. 2020. http://pub1-bjyt.s3.360.cn/bcms/Darkhotel%EF%BC%88APT-C-06%EF%BC%89%E4%BD%BF%E7%94%A8%E2%80%9C%E5%8F%8C%E6%98%9F%E2%80%9D0Day%E6%BC%8F%E6%B4%9E%EF%BC%88CVE-2019-17026%E3%80%81CVE-2020-0674%EF%BC%89%E9%92%88%E5%AF%B9%E4%B8%AD%E5%9B%BD%E5%8F%91%E8%B5%B7%E7%9A%84APT%E6%94%BB%E5%87%BB%E5%88%86%E6%9E%90.pdf

[49] 汪嘉来, 张超, 戚旭衍, 荣易. Windows 平台恶意软件智能检测综述. 计算机研究与发展, 2021, 58(5): 977–994. [doi: 10.7544/issn1000-1239.2021.20200964]

[62] 肖达, 刘博寒, 崔宝江, 王晓晨, 张索星. 基于程序基因的恶意程序预测技术. 网络与信息安全学报, 2018, 4(8): 21–30. [doi: 10.11959/j.issn.2096-109x.2018069]

[79] 吕杨琦, 王张宜, 杨秀璋, 宋文纳, 彭国军. 基于特征功能函数的 APT 样本分类方法. 郑州大学学报 (理学版), 2023, 55(2): 10–17, 24. [doi: 10.13705/j.issn.1671-6841.2021417]

[80] 钱雨村, 彭国军, 王滢, 梁玉. 恶意代码同源性分析及家族聚类. 计算机工程与应用, 2015, 51(18): 76–81. [doi: 10.3778/j.issn.1002-8331.1411-0342]

[81] 安天安全研究与应急处理中心. 白象的舞步——来自南亚次大陆的网络攻击. 2016. https://www.antiy.cn/research/notice&report/research_report/304.html

[93] 赵晶晶. Windows 平台恶意软件的作者组织特征研究 [硕士学位论文]. 广州: 暨南大学, 2020. [doi: 10.27167/d.cnki.gjinu.2020.000381]

[97] 李腾, 乔伟, 张嘉伟, 高怿旸, 王申奥, 沈玉龙, 马建峰. 隐私保护的基于图卷积神经网络的攻击溯源方法. 计算机研究与发展, 2021, 58(5): 1006–1020. [doi: 10.7544/issn1000-1239.2021.20200942]

[108] 乔延臣, 云晓春, 张永铮, 李书豪. 基于调用习惯的恶意代码自动化同源判定方法. 电子学报, 2016, 44(10): 2410–2414. [doi: 10.3969/j.issn.0372-2112.2016.10.019]

[110] 吴鹏. 多形态软件代码同源判定技术研究 [博士学位论文]. 成都: 四川大学, 2021. [doi: 10.27342/d.cnki.gscdu.2021.000007]

[111] 王梓晗. 勒索软件追踪溯源技术研究 [硕士学位论文]. 北京: 北京邮电大学, 2019.

[113] 鲁宁, 张嘉伟, 马建峰, 丛鑫, 史闻博, 王尚广. 可扩展性增强的动态确定包标记溯源方法. 计算机学报, 2020, 43(8): 1493–1516. [doi: 10.11897/SP.J.1016.2020.01493]

[115] 徐恪, 朱亮, 朱敏. 互联网地址安全体系与关键技术. 软件学报, 2014, 25(1): 78–97. http://www.jos.org.cn/1000-9825/4509.htm [doi: 10.13328/j.cnki.jos.004509]

[117] 鲁宁, 张俊伟, 马建峰, 程庆丰, 张嘉伟, 王尚广. 联盟模式下高效单包溯源方法研究. 软件学报, 2020, 31(12): 3880–3908. http://www.jos.org.cn/1000-9825/5882.htm [doi: 10.13328/j.cnki.jos.005882]

[118] 姜建国, 王继志, 孔斌, 胡波, 刘吉强. 网络攻击源追踪技术研究综述. 信息安全学报, 2018, 3(1): 111–131. [doi: 10.19363/j.cnki.cn10-1380/tn.2018.01.008]

[127] 安天. 安天捕风蜜罐系统. 2022. https://www.antiy.cn/Security_Product/ACS.html

[128] 奇安信. 攻击诱捕系统. 2022. https://www.qianxin.com/product/detail/pid/426

[132] 奇安信. 威胁情报的层次分析. 2016. https://ti.qianxin.com/blog/articles/level-of-threat-intelligence/

[133] 冷涛, 蔡利君, 于爱民, 朱子元, 马建刚, 李超飞, 牛瑞丞, 孟丹. 基于系统溯源图的威胁发现与取证分析综述. 通信学报, 2022, 43(7): 172–188. [doi: 10.11959/j.issn.1000-436x.2022105]

[136] 张杰, 彭国军, 杨秀璋. 基于动态 API 调用序列和机器学习的恶意逃避样本检测方法. 山东大学学报 (理学版), 2022, 57(7): 85–93, 102. [doi: 10.6040/j.issn.1671-9352.2.2021.117]

[146] 王永庆. 人工智能原理与方法. 西安: 西安交通大学出版社, 1998.

[151] 张书钦, 白光耀, 李红, 张敏智. 多源数据融合的物联网安全知识推理方法. 计算机研究与发展, 2022, 59(12): 2735–2749. [doi: 10.7544/issn1000-1239.20210954]

[157] 邢倩倩. 基于智能规划的网络安全风险评估 [硕士学位论文]. 长沙: 国防科学技术大学, 2014.

[164] 杨秀璋, 彭国军, 李子川, 吕杨琦, 刘思德, 李晨光. 基于 BERT 和 BiLSTM-CRF 的 APT 攻击实体识别及对齐研究. 通信学报, 2022, 43(6): 58–70. [doi: 10.11959/j.issn.1000−436x.2022116]

[180] 陈伟翔. 面向 APT 家族分析的攻击路径预测方法研究 [硕士学位论文]. 广州: 广州大学, 2021. [doi: 10.27040/d.cnki.ggzdu.2021.000867]

[181] 张耀方, 张哲宇, 曲海阔, 张格, 王子博, 王佰玲. 面向大规模工控网络的关键路径分析方法. 网络与信息安全学报, 2021, 7(6): 31–43. [doi: 10.11959/j.issn.2096-109x.2021069]

[185] 胡浩, 刘玉岭, 张红旗, 杨英杰, 叶润国. 基于吸收 Markov 链的网络入侵路径预测方法. 计算机研究与发展, 2018, 55(4): 831–845. [doi: 10.7544/issn1000-1239.2018.20170087]

[186] 余洋, 夏春和, 胡潇云. 采用混和路径攻击图的防御方案生成方法. 浙江大学学报 (工学版), 2017, 51(9): 1745–1759. [doi:

10.3785/j.issn.1008-973X.2017.09.009]
[188] 翟海霞, 卢月萌, 王辉, 敖山. 基于 T_NAG 模型的攻击路径预测方法研究. 计算机应用研究, 2021, 38(3): 886–892. [doi: 10.19734/j.issn.1001-3695.2019.11.0701]
[189] 孙澄, 胡浩, 杨英杰, 张红旗. 基于网络防御知识图谱的 0day 攻击路径预测方法. 网络与信息安全学报, 2022, 8(1): 151–166. [doi: 10.11959/j.issn.2096-109x.2021101]
[191] 卢嘉中. 基于全网流量与日志深度分析的 APT 攻击建模与检测技术 [博士学位论文]. 成都: 电子科技大学, 2019.
[199] 马琳茹, 杨林, 王建新, 唐鑫. 利用模糊聚类实现入侵检测告警关联图的重构. 通信学报, 2006, 27(9): 47–52. [doi: 10.3321/j.issn:1000-436X.2006.09.008]
[200] 伏晓, 石进, 谢立. 用于自动证据分析的层次化入侵场景重构方法. 软件学报, 2011, 22(5): 996–1008. http://www.jos.org.cn/1000-9825/3759.htm [doi: 10.3724/SP.J.1001.2011.03759]
[203] 王硕, 汤光明, 王建华, 孙怡峰, 寇广. 基于因果知识网络的攻击场景构建方法. 计算机研究与发展, 2018, 55(12): 2620–2636. [doi: 10.7544/issn1000-1239.2018.20160940]
[207] 琚安康. 基于多源异构数据的定向网络攻击检测关键技术研究 [博士学位论文]. 郑州: 战略支援部队信息工程大学, 2020.
[216] 贾召鹏, 方滨兴, 刘潮歌, 刘奇旭, 林建宝. 网络欺骗技术综述. 通信学报, 2017, 38(12): 128–143. [doi: 10.11959/j.issn.1000-436x.2017281]
[219] 杨俊岭, 周宇, 周嘉申. 美军智能欺骗与反欺骗技术发展. 国防科技, 2022, 43(4): 48–53. [doi: 10.13943/j.issn1671-4547.2022.04.09]
[220] 诸葛建伟, 唐勇, 韩心慧, 段海新. 蜜罐技术研究与应用进展. 软件学报, 2013, 24(4): 825–842. [doi: 10.3724/SP.J.1001.2013.04369]
[232] 陈晋音, 胡书隆, 邢长友, 张国敏. 面向智能渗透攻击的欺骗防御方法. 通信学报, 2022, 43(10): 106–120. [doi: 10.11959/j.issn.1000−436x.2022202]
[234] 宋宇波, 樊明, 杨俊杰, 胡爱群. 一种基于拓扑分析的网络攻击流量分流和阻断方法. 信息网络安全, 2020, 20(3): 9–17. [doi: 10.3969/j.issn.1671-1122.2020.03.002]
[246] 纪守领, 杜天宇, 李进锋, 沈超, 李博. 机器学习模型安全与隐私研究综述. 软件学报, 2021, 32(1): 41–67. http://www.jos.org.cn/1000-9825/6131.htm [doi: 10.13328/j.cnki.jos.006131]

杨秀璋(1991－), 男, 博士, CCF 学生会员, 主要研究领域为网络与信息系统安全, 恶意代码检测.

田杨(1996－), 男, 硕士生, 主要研究领域为网络与信息系统安全.

彭国军(1979－), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为网络与信息系统安全, 恶意代码检测.

李晨光(1999－), 男, 硕士生, 主要研究领域为网络与信息系统安全.

刘思德(1997－), 男, 博士生, CCF 学生会员, 主要研究领域为恶意代码检测, 系统安全.

傅建明(1969－), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为系统安全, 网络安全.