基于溯源图分析的高级持续威胁检测技术综述

张　奘[1,2]，杨晓帆[1,2]
（1. 中广电广播电影电视设计研究院有限公司，北京　100045；
2. 广播电视与视听新媒体智慧监管国家广播电视总局重点实验室，北京　100045）

摘　要：全球高级持续威胁（Advanced Persistent Threat，APT）以其高度组织化、隐蔽性强、长期潜伏和跨平台协同的复杂攻击模式，对国家网络空间安全提出严峻挑战。在APT攻击过程中，溯源图能够通过多源数据融合与图结构分析，有效捕获攻击者遗留的蛛丝马迹，对APT攻击检测具有重要帮助。聚焦基于溯源图分析的APT检测技术，对近期国际高水平期刊和会议的工作进行了总结。首先对APT定义、生命周期及当前我国面临的APT攻击现状进行描述；随后，将基于溯源图分析的APT攻击检测方法分为基于传统技术和基于学习训练的方法进行具体介绍和总结，对比优势和局限性，并总结和讨论该领域未来展望，指出将传统方法与学习模型融合研究是未来重要方向，为该领域人员提供借鉴和参考。

关键词：高级持续威胁；溯源图分析；检测；网络安全
中图分类号：TP399；TP309　　**文献标识码：**A　　**DOI：**10.19358/j.issn.2097－1788.2025.08.001
引用格式：张奘，杨晓帆. 基于溯源图分析的高级持续威胁检测技术综述[J]. 网络安全与数据治理，2025，44(8)：1－9.

A survey of provenance graph-based methods for advanced persistent threat detection

Zhang Yan[1,2]，Yang Xiaofan[1,2]
（1. Radio，Film and Television Design and Research Institute Co.，Ltd.，Beijing 100045，China；2. Key Laboratory of Intelligent Supervision for Radio，Television and Audiovisual New Media，National Radio and Television Administration，Beijing 100045，China）

Abstract：Global Advanced Persistent Threats（APTs）pose severe challenges to national cyberspace security due to their highly organized，stealthy，persistent，and cross-platform coordinated attack patterns. During APT attacks，provenance graphs constructed through multi-source data can effectively capture the traces left by attackers，thereby playing a critical role in APT detection. This paper focuses on provenance graph-based APT detection methods and systematically summarizes recent studies from international journals and conferences. Firstly，it delineates the definition of APTs，their lifecycle，and the current landscape of APT attacks faced by China. Subsequently，it categorizes and elaborates on provenance graph-based APT detection methods，dividing them into traditional technique-based methods and learning-based methods. The paper compares their advantages and limitations，summarizes and discusses future prospects in this field，and highlights that integrating traditional methods with learning models represents a critical research direction. This research provides reference guidance for researchers in this field.

Key words：advanced persistent threat；provenance graph analysis；detection；cyber space security

0　引言

随着互联网技术的高速发展和数字化进程的加速推进，网络空间与人类活动息息相关，已成为国家主权、经济命脉和社会稳定的战略要地。然而，高级持续威胁（Advanced Persistent Threat，APT）的泛滥已成为现代计算环境安全面临的最大威胁之一。例如，具有美国背景的APT组织开发的震网（Stuxnet）蠕虫病毒，高度精准破坏伊朗的核计划[1]；“海莲花”（OceanLotus）组织长期利用网络设备漏洞对我国政府、科研院校等高价值目标展开攻击；“方程式”（Equation）组织针对我国西北工业大学实施APT攻击[2]。这类攻击以高度组织化、隐蔽性强、持续周期长为特征，通过多阶段渗透和复杂手段突

破目标网络防御体系，对关键基础设施、核心数据资产乃至国家安全构成严峻挑战。在此背景下，如何实现APT攻击的精准检测与有效防御，成为网络安全领域亟待突破的核心难题。传统安全防护技术（如入侵检测系统等）主要依赖已知攻击特征码或行为模式的匹配，在应对多阶段高隐蔽的APT攻击时存在明显局限性。APT攻击通常采用定向“钓鱼”、定制化漏洞工具、供应链渗透及多跳横向移动等多种策略，其攻击链的碎片化、低频化特征使得基于单点日志或流量分析的方法难以捕捉全局威胁，亟须探索能够深度挖掘攻击上下文关联、适应动态威胁环境的检测范式。

近年来，溯源图分析技术因其关联强覆盖广的优势逐渐成为APT检测领域的研究热点。该技术通过构建实体间关联有向图模型，将离散的系统事件（如进程创建、文件访问、网络连接）映射为具有时序逻辑的语义丰富的结构化数据图表示。通过揭示攻击者从初始入侵到横向渗透的多步行为上下文，溯源图分析技术能够将因果相关的攻击事件通过上下文序列直接关联，有效提升APT攻击检测的准确率。

本文主要对近10年在CCF推荐国际学术高水平（CCF A类）会议和期刊中发表的APT检测相关工作进行深入调研，给出高级持续威胁的定义，重点介绍和分析基于溯源图分析技术的APT检测工作，并讨论了基于溯源图分析技术的APT检测未来发展的侧重点。

1　高级持续威胁

1.1　高级持续威胁定义

高级持续威胁（Advanced Persistent Threat，APT）术语由2006年美国空军上校Gregory Rattray首次提出，聚焦有目的性、有针对性的渗透活动。

随着对APT认识不断深入，美国国家标准与技术研究院（National Institute of Standards and Technology，NIST）于2011年对APT提出了较为具体形象定义：高级持续威胁是指具备高度专业知识和资源的对手（如国家支持的黑客组织或组织化犯罪集团）发起的具有高度针对性和隐蔽性，通过长时期逐步实现窃取敏感数据或破坏关键任务的战略目标。该定义精准地概括了高级持续威胁的核心特征，既强调了其技术复杂性和攻击者的高度组织性，又突出了其长期潜伏、定向渗透的核心策略，并区分了APT与传统威胁的差异。

1.2　高级持续威胁生命周期

根据文献［3］对APT杀伤链的描述，如图1所示，APT攻击生命周期具体可分为：初步侦察、初次渗透、建立据点、权限提升、内网侦察、横向移动、持续化维持、目标达成。

初步侦察：此阶段攻击者会通过多种手段收集目标的基本信息，例如使用开源情报、网络测绘扫描工具等，收集目标的IP地址范围、域名、员工邮箱地址、社交媒体账号、供应商关系等信息，为后续攻击提供充分数据准备。

初次渗透：攻击者利用收集到的信息，寻找目标网络边界的薄弱点进行尝试入侵。常见的入侵手段包括漏洞攻击、钓鱼邮件以及供应链攻击等，以达到入侵目标系统的目的。

建立据点：成功入侵目标网络边界系统后，攻击者会在目标系统中建立据点巩固访问权限，通常会安装后门程序、创建隐藏用户账户、修改系统配置，以防止被发现并确保持续访问。

权限提升：为了能够访问更敏感的数据和系统资源，攻击者通过利用本地漏洞或提权工具，获取更高权限，如从普通用户权限提升到管理员权限或域管理员权限。

内网侦察：此阶段攻击者通过网络边界向目标网络内部进行侦察，以了解网络拓扑结构、关键系统位置，收集网络设备和服务器的配置信息，识别运行的关键服务和应用程序；通过扫描内部网络，发现其他主机和网络设备的IP地址、开放端口和服务等，以便进一步扩展访问权限。

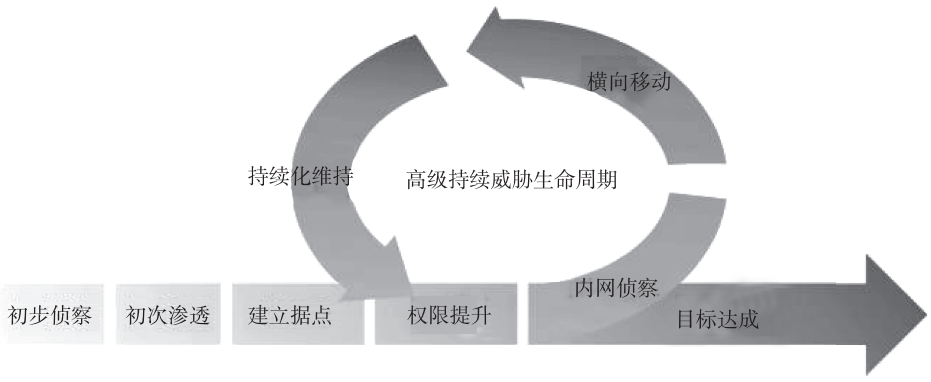

图1　高级持续威胁生命周期图

横向移动：攻击者利用掌握的内网情报，在目标网络中扩散，控制更多主机和系统，以接近高价值目标数据，常见的横向移动方法包括漏洞攻击、口令猜测和伪造凭证等手段。

持续化维持：攻击者会采取多种措施确保其在目标系统中的长期存在，并尽量避免被发现，会定期更新恶意软件或后门程序，以躲避安全检测工具的查杀；使用 Rootkit 或 Bootkit 等技术隐藏其进程、文件和网络活动；篡改系统日志或使用日志擦除工具，以消除其入侵痕迹。

目标达成：在完成数据收集或破坏后，攻击者会将窃取的数据通过隐蔽通道传输回其控制的服务器，传输过程中可能会使用加密、压缩或分段传输等技术，以避免被检测到。最后，攻击者可能会清除在目标系统中留下的痕迹，如删除恶意软件、关闭后门、恢复被篡改的系统配置等，以降低被发现的风险，但也可能留下部分后门以备后续再次入侵。

1.3 我国面临高级持续威胁情况

根据 APT 分析报告[1,4-5]分析，我国长期受到来自美国、印度等国家面向政府、军事、外交、科研、航空航天、核研究、通信和金融等多行业领域的 APT 攻击。表 1 展示了部分针对我国的 APT 组织信息。

2 基于溯源图分析的 APT 检测技术

尽管 APT 攻击手段具有高度隐蔽性，攻击者可能通过伪装绕过传统的单点检测技术，但 APT 攻击在其从侦察到数据窃取或实施破坏整个生命周期中，难免会留下可被识别的踪迹。因此，将 APT 攻击的完整生命周期纳入检测范围，对提高检测准确率具有重要帮助。近年来，基于溯源图的 APT 攻击检测方法因其能够有效关联和可视化攻击事件，逐渐成为学界和企业研究的热点[6-7]，并发展成为主流的 APT 攻击检测方法之一。本节将系统介绍基于溯源图分析的 APT 攻击检测技术，包括溯源图的定义与构建、基于传统技术的 APT 攻击检测方法和基于学习的 APT 攻击检测方法。表 2 列举了部分基于溯源图分析的 APT 检测方法。

2.1 溯源图的定义和构建

溯源图是一种以图形化方式对 APT 攻击过程中涉及的实体及其相互关系进行表示的结构，如图 2 所示。它将网络中的各种对象（如主机、用户、进程、文件、网络连接等）作为节点，将这些对象之间的交互行为，例如进程间的通信、文件访问和读写以及网络数据传输等，作为边进行连接，从而形成一个能够全面反映 APT 各生命周期中事件发生时网络状态和攻击行为传播路径的图结构。图 2 展示的溯源图描述了恶意脚本“/tmp/update.sh”被用户执行后，从恶意域名下载木马并执行后写入定时任务，读取 SSH 私钥以及与 C2 服务器回连的过程。通过对溯源图的分析，能够识别系统当前被攻击的状态，追踪攻击的源头，了解攻击的传播过程以及发现攻击过程中潜在的恶意活动和关键节点，为 APT 攻击的检测和防御提供有力支持。

表 1 部分针对我国的 APT 组织信息

序号	APT 组织名称	组织编号	地理位置	最早发现时间	攻击行业领域
1	Longhorn	APT-C-39	美国	2008 年	航天、科研、能源
2	TaskMasters	无	英语地区	2010 年	政府、能源、工业、技术
3	蓝宝菇	APT-C-12	东亚地区	2011 年	政府、军工、核研究
4	索伦之眼	APT-C-16	美国	2011 年	政府、通信、金融
5	摩诃草	APT-C-09	印度洋地区	2013 年	政府、科研、教育
6	飞鲨	APT-C-17	印度洋地区	2013 年	航空航天
7	方程式	APT-C-40	美国	2015 年	政府、航天、核研究、军事
8	海莲花	APT-C-00	越南	2016 年	政府
9	CozyBear	APT 29	东欧地区	2016 年	政府
10	蔓灵花	APT-C-08	印度	2017 年	政府、电力、工业
11	响尾蛇	APT-C-24	印度洋地区	2017 年	政府、外交、军事
12	Emissary Panda	APT 27	未知	2017 年	国防、政府、航空航天
13	CNC	APT-C-48	印度洋地区	2019 年	军工、教育、医疗
14	魔罗枞	无	未知	2020 年	政府、军事、核能
15	双异鼠	无	未知	2023 年	政府
16	水粉虫	无	未知	2024 年	通信

表 2　基于溯源图分析的 APT 检测方法部分列表

文献	发表时间	APT 攻击检测方法	测试数据集	是否开源	刊物/会议	开源地址
StreamSpot[8]	2016	基于聚类异常检测（基于学习的方法）	基于 YouTube、CNN、GMail 等模拟数据	否	SIGKDD	未提及
SLEUTH[9]	2017	基于规则策略匹配（基于传统技术）	DARPA 数据集等	否	USENIX Security	未提及
Holmes[3]	2019	基于规则策略匹配和启发式算法（基于传统技术）	DARPA 数据集等	否	IEEE S&P	未提及
Unicorn[10]	2020	聚类算法（基于学习的方法）	DARPA 数据集、StreamSpot 数据集等	是	NDSS	https://github. com/crimson-unicorn
Alchemist[11]	2021	基于 Datalog 的跨日志关系推理（基于传统技术）	DARPA 数据集等	部分开源	NDSS	https://github. com/ALchemist2020/Workload
ProGrapher[12]	2023	基于图嵌入和 extRCNN 序列学习异常检测（基于学习的方法）	StreamSpot 数据集、ATLAS 数据集、DARPA 数据集等	否	USENIX Security	未提及
MAGIC[13]	2024	异常检测和节点异常识别（基于学习的方法）	DARPA 数据集、StreamSpot 数据集、Unicorn 数据集等	是	USENIX Security	https://github. com/MAGIC-APT/MAGIC
Flash[14]	2024	基于图神经网络和 XGBoost 异常检测（基于学习的方法）	DARPA 数据集、StreamSpot 数据集、Unicorn 数据集等	是	IEEE S&P	https://github. com/DART-Laboratory/Flash-IDS
VoDKA[15]	2025	基于知识蒸馏的轻量级框架节点分类（基于学习的方法）	DARPA 数据集、StreamSpot 数据集、Unicorn 数据集等	否	WWW	未提及

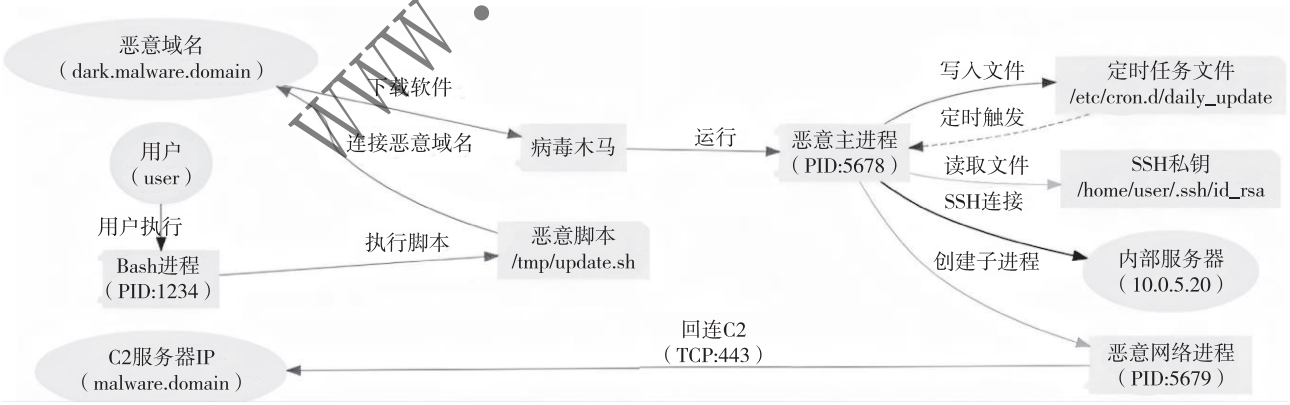

图 2　溯源图示例

溯源图的构建过程可分为 4 个阶段：数据收集、实体识别与提取、关系发现与建立、图结构化表示。

在数据收集阶段，通过收集多种数据源的信息，包括但不限于网络流量数据、主机日志（系统日志、应用程序日志等）、终端监控数据（进程运行状态、文件操作记录、注册表修改记录等）以及外部威胁情报等信息形成溯源图的基础数据。

在实体识别与提取阶段，利用正则表达式或其他识别算法，从收集的数据中识别出实体，如 IP 地址、域名、进程、文件、用户等，用于构建溯源图的节点。

在关系发现与建立阶段，通过时间序列关联分析、行为特征关联分析或其他关联规则挖掘算法，发现实体

间的关联关系，确定关系的方向、强度等属性。例如，行为特征分析确认进程实体读取或写入某本地文件，并生成该关系边。

在图结构化表示阶段，将实体作为节点，关系作为边，按照一定的拓扑结构构建溯源图，选择有向图或无向图表示，为节点和边赋予相应属性，以更全面地描述实体和关系的性质。

溯源图作为一种核心数据结构，通过对网络攻击行为的可视化建模，有效刻画 APT 攻击的路径与实体关联关系，能够为 APT 攻击检测提供较为全面的数据支撑。

2.2 基于传统技术的 APT 攻击检测方法

基于传统技术的 APT 攻击检测方法基于标签和规则策略实现高效、快速且精准完整的 APT 攻击检测[16-17]。但实际中，溯源图往往具备较大规模，这给检测效率带来了巨大挑战。溯源图以系统、进程、文件实体作为节点，因此其增长极为迅速，可能存在大量冗余信息，使得威胁行为隐藏在海量数据之中。研究人员通过为节点添加威胁标签信息来帮助定位 APT 攻击行为，并依据先验知识、专家经验和标签信息为边赋予权重，进而根据标签和权重来进行 APT 攻击检测。

Lee 等人针对在长期运行程序场景下的日志依赖爆炸问题，提出了基于执行单元的细粒度因果依赖图检测方法 BEEP（Binary-based Execution Partition）[18]。该方法通过动态逆向工程识别程序中的事件处理循环，将进程执行划分为多个逻辑独立的执行单元，每个单元对应特定输入请求的处理过程。BEEP 通过二进制插桩技术自动识别此类循环边界，并逆向检测触发单元间依赖的关键内存操作指令，在运行时记录单元划分和跨单元工作流依赖。在因果依赖图构建上，BEEP 将传统进程级审计日志与单元级内存操作日志相结合用于支撑 APT 检测。

Hossain 等人提出的 SLEUTH[9] 构建了基于审计数据的轻量化内存依赖图模型，为解决大规模溯源图分析难题提供了一种新思路。该研究针对传统图存储与计算效率瓶颈，设计了超紧凑的主存数据结构，将平均事件存储密度压缩至 10 字节/边，支持每秒处理超过 20 000 个事件的实时分析能力。其核心创新在于构建多维度标签体系，通过信任度与敏感度双重标签的动态传播机制，实现对攻击行为的精准识别与路径聚焦。研究团队提出基于 Dijkstra 算法的双向加权搜索策略，通过标签引导的前向影响分析与后向溯源分析，将攻击场景重构的计算复杂度从 O（n^2）降至线性级。另外，SLEUTH 创新性地引入可定制的策略框架，允许通过正则表达式灵活定义文件、进程的初始标签与传播规则，既保留默认策略的保守性优势，又支持特定应用场景（如 SSH 服务、软件更新）的误报消除。

Morse[19] 是一种基于动态标签传播的 APT 攻击场景重构方法。它通过创新的标签衰减（Tag Attenuation）和标签退化（Tag Decay）机制应对取证分析中的依赖爆炸问题。该研究针对传统粗粒度溯源分析导致的海量误报问题，构建了一种双维度标签体系：数据标签表征数据的机密性和完整性，主体标签区分进程的可疑等级。通过设计差异化的标签传播规则，系统对良性进程采取宽松传播策略，利用其常规行为模式降低误报；对可疑进程则采用保守传播策略，确保攻击路径完整保留。在实验验证环节，Morse 能成功检测基于内存载荷、浏览器扩展和内核恶意软件等隐蔽攻击。Morse 的核心创新在于建立了基于进程行为特性的动态标签调节机制。标签衰减通过引入线性叠加因子，削弱良性进程通过多跳传播恶意标签的能力。标签退化采用指数衰减模型，逐步恢复长期运行的良性进程标签可信度。

Holmes[3] 是一种基于可疑信息流关联的实时 APT 检测方法，其创新性在于构建了面向战术意图的高层场景图，通过多阶段攻击行为的因果关联实现了对 APT 活动的精准刻画。该方法通过引入 MITRE ATT&CK 框架中的战术技术流程（Tactics, Techniques, and Procedures, TTP）作为中间语义层，构建了从底层系统调用到高层攻击意图的映射机制。它采用三层架构：在数据层构建版本化的紧凑溯源图以支持实时分析；在语义层建立基于 TTP 的战术意图识别机制，将百万级底层事件压缩至百级可疑活动；在检测层创新性地提出威胁元组（Threat Tuple）模型，结合加权乘积公式对攻击阶段进行动态评分，有效区分正常操作与隐蔽攻击。

Hassan 等人针对现有因果分析方法中系统层与应用层语义割裂的核心挑战，提出了多层级日志融合的全局溯源图（Universal Provenance Graph, UPG）构建方法 OmegaLog[20]。该方法表示应用层事件日志在攻击检测中具备重要价值：开发者通过日志语句天然标记的事件处理循环（Event-Handling Loop）不仅包含高价值语义信息，还能为执行单元划分（Execution Partitioning）提供精确边界。OmegaLog 设计了静态二进制分析与动态日志拦截相结合的三阶段框架：通过符号执行与路径分析从程序二进制中提取日志消息字符串（Log Message Strings, LMS）的控制流模型；运行时通过内核模块捕获进程级日志事件并与系统审计日志时空对齐；最终基于 LMS 控制流模式将应用事件无缝嵌入系统层溯源图。该方法在无需开发者干预或训练的前提下，实现了应用层执行单元的细粒度划分，一定程度上缓解了依赖爆炸问题。

综上，传统方法以规则驱动和标签策略为核心，依

赖专家知识构建检测模型，在实时性、可解释性和轻量化方面具有显著优势。例如，SLEUTH[9]通过内存数据结构压缩与动态标签传播实现高效分析，Holmes[3]利用APT攻击生命周期意图映射和预设规则策略实现攻击检测。但是，传统方法的规则构建、标签权重分配以及策略调整高度依赖专家经验，对APT攻击的多样性与演化性的覆盖存在一定难度，并且标签权重分配、策略调整均需专家经验支撑，难以应对海量攻击检测场景。

2.3 基于学习的APT攻击检测方法

相较于依赖专家规则与标签策略的传统检测方法，基于学习的APT攻击检测技术通过数据驱动方式自动归纳行为的内在关联与攻击模式，在应对大规模溯源图分析与复杂攻击场景重构方面表现出优势。传统方法虽能通过标签加权与规则匹配实现高效检测，但其性能受限于专家经验的完备性与规则设计的可扩展性。近年来研究者逐步将深度学习、图表示学习与知识蒸馏等先进技术引入APT检测领域[21-25]，通过构建自动化特征提取与语义建模框架，有效提升了APT攻击检测能力与检测鲁棒性。

Alsaheel等人提出了一种融合因果分析、自然语言处理和序列建模的新型APT攻击检测方法ATLAS[26]，为大规模溯源图的场景重建难题提供了一种解决方案。该研究认为，尽管不同APT攻击的漏洞利用方式存在差异，但其攻击策略在抽象层面具有相似性。基于此观察，ATLAS采用三阶段处理流程：首先通过语义保留的因果图优化算法将原始审计日志压缩为精简的因果依赖图，继而利用词形还原和词嵌入技术将节点行为转化为具有语义表征的时序序列，最终通过双向LSTM模型学习攻击模式与正常行为的区分特征。实验表明，该方法在10个真实APT攻击场景中实现了较高的精确率和召回率。

DepImpact[27]针对大规模溯源图分析效率低下的问题，提出了一种基于多维特征投影的检测方法。DepImpact通过量化系统依赖关系的上下文语义特征，构建动态权重模型以区分攻击路径与正常行为，实现攻击场景的精简重构。该方法首先从时序相关性、数据流相似性和节点集中度三个维度提取边特征，利用线性判别分析对特征空间进行最优投影，生成能够最大化攻击相关边与非关键边区分度的依赖权重。该方法通过局部归一化处理保留了远距离关键边的语义显著性，有效解决了全局权重分配中的梯度消失问题。在权重计算基础上，设计反向依赖影响传播算法，通过迭代式权重聚合量化各节点对事件的影响力，进而识别出关键攻击入口节点。

MAGIC[13]是一种基于掩码图表示学习的自监督APT检测框架，突破了传统方法对攻击先验知识和专家规则的依赖。该方法通过构建精简化的溯源图，并设计了双层图表示学习机制：在节点层面，采用掩码自编码器重构被随机遮蔽的系统实体特征，使得模型深入挖掘实体间的上下文关联；在全局层面，通过采样对比学习捕捉系统行为的拓扑结构特征。这种联合学习策略使得模型能够自动捕获正常系统行为的深层语义模式，而无需依赖人工定义的标签或权重。检测阶段采用基于距离的异常评分机制，通过对比待测实体与良性行为表征的空间偏离度实现多粒度威胁识别，既支持批量日志级别的APT存在性检测，也支持精准定位具体恶意实体。

针对溯源图分析中全局噪声干扰与检测效率协同优化的难题，Wu等人提出了分析工具VoDKA[15]系统，将知识蒸馏框架引入APT攻击检测领域，通过异构模型协同与知识迁移机制，在保持高检测精度的同时显著降低计算负载。VoDKA首先基于图信号处理理论，设计面向系统实体交互的拓扑感知降噪方法，在保留关键攻击路径结构特征的前提下，有效压缩溯源图规模。实验表明，该方法通过动态平衡权重参数实现模型蒸馏的精度可控性，在攻击检测率上较传统图神经网络模型有所提升，时延降低约41.6%。值得注意的是，VoDKA系统引入的个性化PageRank标签传播机制能够有效识别长链路潜伏攻击的弱关联行为特征，解决传统方法因局部泛化能力不足导致的路径截断问题，但在图压缩过程中基于拓扑感知降噪的同时，可能弱化部分低频但关键的异常路径信号而影响性能。相较于依赖专家规则赋权的早期方法，VoDKA框架通过数据驱动方式自适应的知识蒸馏过程，为降低大规模溯源图分析对人工经验的依赖提供了可行的途径。

Han等人提出了一种基于数据溯源图的实时APT检测框架Unicorn[10]，通过融合系统级行为建模与动态图分析技术，有效提升了复杂隐蔽攻击的检测能力。Unicorn针对APT攻击“低频慢速”特征，构建了面向全系统溯源图的动态分析范式，通过捕获进程、文件、网络连接等实体间的因果依赖关系，突破传统系统调用序列分析在上下文关联性和时间跨度上的局限性。基于改进Weisfeiler-Lehman算法[28]的多跳邻域图特征提取机制，通过迭代式标签传播捕捉节点在多阶路径中的上下文语义，来识别跨时间窗口的异常信息流模式。实验评估表明，Unicorn在DARPA透明计算数据集和供应链攻击场景中实现了较高的检测准确率，较StreamSpot方法[8]在误报率和检测延迟上均有所降低。Unicorn验证了全系统溯源图在实时APT检测中的可行性，其轻量级图摘要机制与增量式分析框架为处理长期系统监控产生的大规模事件流提供了一种可行的途径。

Wang 等人提出的 Threatrace[29] 为解决隐蔽攻击场景下的节点级异常检测问题提供了重要思路。文中指出，基于全图特征（如 StreamSpot[8] 和 Unicorn[10]）或路径分析（如 ProvDetector[30]）的方法难以有效捕捉 APT 攻击中高度分散且伪装性强的局部异常行为。为此，Wang 等人将图神经网络与系统溯源图分析相结合，提出一种基于节点上下文语义的实时检测框架。通过改进的 GraphSAGE 模型构建节点级特征表示，利用图结构传播机制捕捉实体在信息流中的动态角色，而非依赖传统路径或全图相似性度量。Threatrace 中提出一种流式处理架构，采用“磁盘－内存”协同存储策略，在保持历史上下文完整性的同时有效降低内存消耗，支持对长期运行系统的持续监控。值得注意的是，Threatrace 能够在无需攻击先验知识情况下进行无监督学习，通过分析节点类型与行为模式的统计偏差定位异常，成功检测出包括提权操作、隐蔽通信在内的 11 类 MITRE ATT&CK 战术。该工作开源在 GitHub 平台上，为后续研究提供了可复现的原型系统。

Flash[14] 是一种基于图神经网络与语义嵌入的溯源图入侵检测方法，旨在解决现有 APT 检测方法在语义信息利用、时序关系建模和系统可扩展性方面的局限。Flash 提出一种双阶段特征编码机制：首先通过改进的 Word2Vec 模型融合进程名称、文件路径等语义属性和事件时序信息，生成具有时间敏感性的节点嵌入；其次设计基于 GraphSAGE 的上下文编码器，通过选择性图遍历策略（如过滤低优先级边、抽象用户特定属性）提取节点局部与全局结构特征。实验评估上，Flash 在 DARPA 数据集上的检测准确率优于 Threatrace、Unicorn 等方法。此外，Flash 具备较强鲁棒性，通过节点级异常检测机制有效抵御了基于邻居结构模仿的对抗攻击。

针对溯源图实时检测场景中动态性与内存效率的挑战，Manzoor 等人提出了一种基于流式聚类的异常检测方法 StreamSpot[8]。该方法面向异构信息流图的实时处理需求，通过融合局部结构特征提取与内存高效表示技术，实现对 APT 攻击行为的低延迟检测。StreamSpot 设计了一种基于 k-shingling 的异构图相似性度量方法，通过有序广度遍历提取节点邻域序列特征，并采用改进的流式哈希（StreamHash）将高维特征映射到固定维度的签名向量。此外，StreamSpot 通过动态维护基于质心的聚类模型，能够在处理无限数据流时控制在常数级内存空间，同时保证对异常图的实时评分能力，将异常检测延迟从传统批处理模式改进为实时边缘处理范式，在保持高精度的同时显著提升了系统吞吐量。在实践中，聚类的半径、特征向量维度仍需根据不同环境进行适应性调整。

Yang 等人引入图嵌入与序列学习技术到基于溯源图的 APT 异常检测工作中，提出 ProGrapher[12] 方法。ProGrapher 针对现有方法在效率与检测粒度上的不足，设计了动态快照构建机制，将连续审计日志划分为时间有序的子图快照，有效缓解了全量溯源图分析时的“依赖爆炸”问题。在特征提取阶段，ProGrapher 采用改进的 graph2vec 算法生成全图嵌入向量，通过最大化正常快照与正常根子图的共现概率，捕捉复杂非线性的图结构特征。为增强时序建模能力，系统引入 TextRCNN 模型[31] 对快照序列进行动态预测，通过对比实际嵌入与预测向量的偏差实现异常检测。实验评估表明，ProGrapher 在 StreamSpot[8]、DARPA 等数据集上均展现出较强性能。该方法支持实时检测，单次推理耗时仅 10 s，内存消耗随数据规模呈次线性增长，具备较好的工程适用性。但 TextRCNN 模型对长时序上下文建模存在局限性，难以捕捉间隔较远的因果依赖，可能导致对长时期潜伏的攻击漏报。

Wang 等人提出的基于溯源数据的 ProvDetector 方法[30]，通过深度挖掘系统行为依赖关系实现了对 APT 攻击的有效检测。针对识别较大的利用合法可信实体执行恶意行为恶意软件，ProvDetector 构建了一种面向进程行为的因果路径分析框架。具体地，ProvDetector 采用基于稀有度的路径选择算法，从海量溯源图中自动筛选出能够表示异常行为的关键因果路径，一定程度上缓解了依赖爆炸问题。通过将路径序列映射为神经网络嵌入表示，并结合局部离群因子密度检测模型，实现了对混合在正常行为中的恶意活动的精准识别，为后续 APT 检测研究提供了可行的技术路径。

Xie 等人提出了一种融合路径分析与图结构特征的混合检测模型 Pagoda[32]。该方法将实时检测与全图分析相结合，采用深度优先搜索对关键攻击路径进行优先级判定，通过路径异常度阈值实现早期攻击识别。若路径检测未触发警报，Pagoda 引入路径长度加权机制对全图异常度进行概率聚合，捕捉分布式攻击特征。针对大规模溯源图处理瓶颈，Pagoda 提出基于内存数据库的轻量化存储架构，通过键值聚合技术将路径依赖关系压缩为单次查询操作，结合字典编码对重复路径模式进行语义压缩，提升规则库存储效率。此外，Pagoda 采用双层过滤机制，在数据采集层主动过滤掉管道文件、临时文件等噪声节点，在分析层忽略环境变量等非关键元数据，构建出面向攻击行为表征的精简后的溯源子图。

综上，基于学习的 APT 检测方法通过海量数据驱动实现自动化特征提取与攻击行为识别，在复杂攻击场景建模与未知威胁检测中表现突出。例如，MAGIC[13] 利用自监督图表示学习捕捉正常行为模式，VoDKA[15] 通过知识蒸馏平衡检测效率与精度，Threatrace[29] 结合图神经网

络与流式处理实现无监督异常检测。基于学习的 APT 检测方法自适应力强，通过训练数据可捕获 APT 攻击的隐蔽行为特征，但是由于需要大量训练数据支撑，因此标注数据质量与覆盖范围对其性能有较大影响。并且基于学习的检测方法需要较大的计算开销，边缘设备部署和实时流处理也存在一定难度。

3 结论

高级持续威胁的泛滥已成为当前网络空间安全的严峻挑战，其多阶段隐蔽渗透、跨平台横向移动及长期潜伏特征，对关键基础设施和国家安全构成系统性风险。APT 攻击的精准检测与防御，是当前核心网络安全难题。APT 攻击检测面临多维度挑战：APT 攻击的隐蔽性、低频性、跨平台协同性以及长期潜伏性对方法的准确率提出了挑战；涉及的实体规模膨胀与依赖爆炸问题，对方法的优化和效率提出了挑战；攻击者采用多种攻击手段，借助正常程序掩盖攻击行为，对方法的鲁棒性提出了挑战。将 APT 攻击的完整生命周期纳入检测范围构建溯源图，对提高检测准确率具有重要帮助。本文对近年来基于溯源图的 APT 攻击检测方法进行了总结归纳，并对基于传统技术的检测方法和基于学习的检测方法的优势和局限性进行了分析。

未来可研究传统方法与基于学习的方法之间的融合演进，具体将传统规则算法与深度学习模型进行互补式集成，在实时 APT 检测系统中实现协同作用；同时，针对多模态数据融合，覆盖系统审计日志、网络流量和应用语义数据源，以实际提升检测性能。传统 APT 攻击检测方法在实时性、可解释性和轻量化方面具有优势，而基于学习的 APT 检测方法在复杂攻击（如多阶段 APT 攻击）场景建模与未知威胁检测中表现突出。可通过对基于学习的 APT 攻击检测方法进行蒸馏，提取生成可用于高效匹配的规则化策略，同时基于传统方法的先验知识和规则的预处理也能有效降低溯源图的规模，提高溯源图抽象表示和异常检测的效率。在多模态数据融合方面，结合系统审计日志、网络流量、应用语义等多源数据，通过深度学习模型（如 BERT、LSTM 等）映射到隐空间进行对齐，利用注意力机制构建多角度行为模型，从多个维度捕获 APT 攻击行为痕迹，提高方法的检测能力。

参考文献

[1] 中国网络安全产业联盟. 美国情报机构网络攻击回顾 [EB/OL]. (2023 - 04 - 11) [2025 - 02 - 18]. https: //chinacia. org. cn/AQLMWebManage/ Resources/kindeditor/attached/file/20230411/20230411161510_ 6312. pdf.

[2] 360 数字安全集团. 美相关 APT 组织分析报告 [EB/OL]. (2024 - 02 - 05) [2025 - 02 - 19]. https: //cdn. isc. 360. com/iscvideo-bucket/APT_ organization_ analysis. pdf.

[3] MILAJERDI S M, GJOMEMO R, ESHETE B, et al. Holmes: real-time APT detection through correlation of suspicious information flows [C]//2019 IEEE Symposium on Security and Privacy (SP). IEEE, 2019.

[4] 360 数字安全集团. 360APT 全景雷达 [EB/OL]. [2025 - 02 - 19]. https: //apt. 360. net/aptlist.

[5] 绿盟科技. APT 组织情报研究年鉴 [EB/OL]. [2025 - 01 - 07]. https: //www. nsfocus. com. cn/html/2022/92_0105/167. html.

[6] 王郅伟, 何睎杰, 易鑫, 等. 基于 APT 活动全生命周期的攻击与检测综述 [J]. 通信学报, 2024, 45 (9): 206 - 228.

[7] ALHANAHNAH M, MA S, GEHANI A, et al. autoMPI: automated multiple perspective attack investigation with semantics aware execution partitioning [J]. IEEE Transactions on Software Engineering, 2023 (4): 49.

[8] MANZOOR E, MILAJERDI S M, AKOGLU L. Fast memory-efficient anomaly detection in streaming heterogeneous graphs [C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1035 - 1044.

[9] HOSSAIN M N, MILAJERDI S M, WANG J, et al. SLEUTH: real-time attack scenario reconstruction from COTS audit data [C]// Proceedings of the 26th USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2017: 487 - 504.

[10] HAN X Y, PASQUIER T, BATES A, et al. Unicorn: runtime provenance-based detector for advanced persistent threats [C]// Network and Distributed Systems Security Symposium 2020. Reston: Internet Society, 2020: 24046.

[11] YU L, MA S, ZHANG Z, et al. ALchemist: fusing application and audit logs for precise attack provenance without instrumentation [C]//28th Annual Network and Distributed System Security Symposium, 2021: 1 - 18.

[12] YANG F, XU J C, XIONG C L, et al. ProGrapher: an anomaly detection system based on provenance graph embedding [C]// Proceedings of the 32nd USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2023: 4355 - 4372.

[13] JIA Z, XIONG Y, NAN Y, et al. MAGIC: detecting advanced persistent threats via masked graph representation learning [C]// Proceedings of the 33rd USENIX Conference on Security Symposium, 2024: 3005 - 3022.

[14] REHMAN M U, AHMADI H, HASSAN W U. Flash: a comprehensive approach to intrusion detection via provenance graph representation learning [C]//2024 IEEE Symposium on Security and Privacy (SP). IEEE, 2024: 3552 - 3570.

[15] WU W, QIAO W, YAN W, et al. Brewing VoDKA: distilling pure knowledge for lightweight threat detection in audit logs [C]//Proceedings of the ACM on Web Conference 2025, 2025: 2172 - 2182.

[16] MA S Q, ZHANG X, XU D. Protracer: towards practical provenance tracing by alternating between logging and tainting [C]// 23rd Annual Network and Distributed System Security Symposium, 2016: 1 - 15.

[17] KWON Y. MCI: modeling-based causality inference in audit logging for attack investigation [C]//25th Annual Network and Distributed System Security Symposium, 2018: 1 - 15.

[18] LEE K H, ZHANG X, XU D. High accuracy attack provenance via vinary-based execution partition [C]//Network and Distributed System Security Symposium, 2013: 16.

[19] HOSSAIN M N, SHEIKHI S, SEKAR R. Combating dependence explosion in forensic analysis using alternative tag propagation semantics [C]//2020 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE, 2020: 1139 - 1155.

[20] HASSAN W U, NOUREDDINE M A, DATTA P, et al. OmegaLog: high-fidelity attack investigation via transparent multi-layer log analysis [C]//Proceedings 2020 Network and Distributed System Security Symposium, 2020.

[21] VAN EDE T, AGHAKHANI H, SPAHN N, et al. DEEPCASE: semi-supervised contextual analysis of security events [C]// 2022 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE, 2022: 522 - 539.

[22] PEI K, GU Z, SALTAFORMAGGIO B, et al. HERCULE: attack story reconstruction via community discovery on correlated log graph [C]//Proceedings of the 32nd Annual Conference on Computer Security Applications. New York: ACM, 2016: 583 - 595.

[23] XU Z Q, FANG P C, LIU C L, et al. DEPCOMM: graph summarization on system audit logs for attack investigation [C]// 2022 IEEE Symposium on Security and Privacy (SP), 2022: 540 - 557.

[24] ZENG J, CHUA Z L, CHEN Y, et al. WATSON: abstracting behaviors from audit logs via aggregation of contextual semantics [C]//28th Annual Network and Distributed System Security Symposium, 2021: 1 - 18.

[25] CHENG Z, LV Q, LIANG J, et al. Kairos: practical intrusion detection and investigation using whole-system provenance [C]// 2024 IEEE Symposium on Security and Privacy (SP). IEEE, 2024: 3533 - 3551.

[26] ALSAHEEL A, NAN Y, MA S, et al. ATLAS: a sequence-based learning approach for attack investigation [C]//Proceedings of the 30th USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2021: 3005 - 3022.

[27] FANG P C, GAO P, LIU C L, et al. Back-propagating system dependency impact for attack investigation [C]//Proceedings of the 31st USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2022: 2461 - 2478.

[28] SHERVASHIDZE N, SCHWEITZER P, LEEUWEN E J V, et al. Weisfeiler-lehman graph kernels [J]. The Journal of Machine Learning Research, 2011, 12 (3): 2539 - 2561.

[29] WANG S, WANG Z, ZHOU T, et al. Threatrace: detecting and tracing host-based threats in node level through provenance graph learning [J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 3972 - 3987.

[30] WANG Q, HASSAN W U, LI D, et al. You are what you do: hunting stealthy malware via data provenance analysis [C]// 27th Annual Network and Distributed System Security Symposium. Reston: Internet Society, 2020: 1 - 17.

[31] LAI S, XU L, LIU K, et al. Recurrent convolutional neural networks for text classification [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2015, 29 (1).

[32] XIE Y L, FENG D, HU Y C, et al. Pagoda: a hybrid approach to enable efficient real-time provenance based intrusion detection in big data environments [J]. IEEE Transactions on Dependable and Secure Computing, 2020, 17 (6): 1283 - 1296.

(收稿日期: 2025 - 06 - 09)

作者简介:

张奕 (1988 -), 男, 硕士, 助理工程师, 主要研究方向: 智慧广电、网络安全。

杨晓帆 (1984 -), 男, 硕士, 高级工程师, 主要研究方向: 智慧广电、网络安全。

版权声明

凡《网络安全与数据治理》录用的文章，如作者没有关于汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权等版权的特殊声明，即视作该文章署名作者同意将该文章的汇编权、翻译权、印刷权及电子版的复制权、信息网络传播权与发行权授予本刊，本刊有权授权本刊合作数据库、合作媒体等合作伙伴使用。同时，本刊支付的稿酬已包含上述使用的费用，特此声明。

《网络安全与数据治理》编辑部