智能溯源分析与入侵检测:洞察、挑战与展望

李振源[1),2)] 韦洋洋[2)] 王征凯[2)] 纪守领[1)]

[1)](浙江大学计算机科学与技术学院 杭州 310007)

[2)](浙江大学软件学院 浙江 宁波 315048)

摘 要 “构建系统行为全局透明可观测性,并通过全局行为关联分析检测攻击”的研究思路在网络空间攻击检测和威胁分析领域受到学术界和产业界的广泛关注和认可。研究者提出利用统一的数据模型对系统行为进行建模,并在此基础上进行攻击检测的方案。方案基于数据流和控制流分析,将系统中的实体和行为抽象为节点和边,并构建成图的形式。由于该图的构造与分析过程类似于数据分析领域的溯源分析(Provenance Analysis),因此被称为溯源图。近年来,基于溯源图分析的检测、分析和取证研究受到全球一流科研机构和大型企业顶尖研发团队的广泛关注。相关成果发表了大量高水平论文。基于这些成果构建的工具与系统在国内外信息技术与网络安全公司得到广泛应用。随着人工智能领域的快速发展,时间序列分析、图表示学习、学习索引、深度图搜索等机器学习技术被大量引入基于溯源的入侵检测系统设计。学术界和业界提出了许多新技术方案,从数据压缩、存储、管理到攻击检测、溯源和重建,对基于溯源的入侵检测系统进行了重新思考与重构,为该领域带来了新的机遇。然而,相应的风险和挑战也随之而来。一方面,网络安全是对抗性学科,而机器学习模型的有效性往往依赖于训练数据的完整性和检测数据的纯净性,容易受到模型投毒、数据污染等攻击,导致模型失效。另一方面,机器学习高昂的训练和预测成本,以及较差的结果可解释性,使其在实际部署时面临计算成本和人工处置成本的压力,影响其实用性。围绕上述问题,本文对近年来基于溯源的入侵检测中机器学习技术应用的相关研究进行了广泛调研和内容梳理,对方法准确率、效率、鲁棒性、结果可解释性等影响系统效果的属性进行了系统性比较,总结了研究现状和未解决的挑战,旨在为该领域未来的研究与应用提供理论支持和灵感。

关键词 入侵检测;威胁狩猎;审计日志分析;溯源分析;图分析;机器学习

中图法分类号 TP309 **DOI号** 10.11897/SP.J.1016.2025.02406

Learning in Provenance-Based Intrusion Detection: A Survey

LI Zhen-Yuan[1),2)] WEI Yang-Yang[2)] WANG Zheng-Kai[2)] JI Shou-Ling[1)]

[1)](College of Computer Science and Technology, Zhejiang University, Hangzhou 310007)

[2)](School of Software Technology, Zhejiang University, Ningbo, Zhejiang 315048)

Abstract In recent years, as cyber threats grow more sophisticated, traditional intrusion detection methods that rely on isolated event signatures have proven insufficient. Consequently, both the cyber security academic and industrial communities seek to enhance system behavior observability and attack detection through global behavior analysis. This initiative advocates for a unified data model to represent system behaviors, developing attack detection schemes based on this model. Such models, which rely on data and control flows, abstract system entities (e.g., processes, files, network sockets) and actions (e.g., read, write, execute) into nodes and edges, forming a graph akin to the provenance graphs used in database fields. This causal dependency

收稿日期:2025-02-08;在线发布日期:2025-07-09。本课题得到浙江省“尖兵领雁”研发计划(Nos. 2025C02263, 2024C03288)、国家自然科学基金青年科学基金项目(No. 62402419)、国家重点研发计划(No. 2023YFB3106800)、宁波市“甬江”人才工程、CCF-绿盟“鲲鹏”科研基金等项目资助。 **李振源**,博士,特聘研究员,中国计算机学会(CCF)会员,主要研究领域为系统安全、网络威胁狩猎。E-mail:lizhenyuan@zju.edu.cn。 **韦洋洋**,硕士研究生,主要研究领域为溯源入侵检测、高级持续性威胁防御。 **王征凯**,硕士研究生,主要研究领域为溯源入侵检测、高级持续性威胁防御。 **纪守领**(通信作者),博士,教授,长江学者,中国计算机学会(CCF)高级会员,主要研究领域为数据驱动安全、AI安全。E-mail:sji@zju.edu.cn。

graph provides a high-fidelity audit trail of system execution. This process of graph construction and analysis, often termed provenance graph analysis, has grown into one of the most prominent methods in intrusion detection. It has garnered widespread attention and recognition from both the academic community and industry, evidenced by numerous publications at top security conferences and in high-level journals. Notably, related methods and technologies have been experimentally adopted by global IT and cybersecurity companies. Recent advancements in artificial intelligence have seen the integration of machine learning techniques, such as time-series analysis, graph representation learning, learning indexes, and deep graph search, into the design of provenance-based intrusion detection systems. These innovations have prompted both academic and industrial communities to propose new schemes, rethinking and reconstructing provenance-based intrusion detection from multiple aspects including data storage, management, compression, attack detection, provenance, and reconstruction. Given that provenance graphs can scale to billions of nodes, intelligent data reduction and efficient querying have become paramount. This evolution presents new opportunities and prospects for the field. However, these advancements also introduce significant risks and challenges. Cyber security is inherently adversarial, and the opaque, data-driven nature of many ML models creates a new attack surface. The effectiveness of machine learning models heavily depends on the integrity of the training data and the purity of the testing data. This vulnerability leaves systems susceptible to model poisoning, data contamination, and carefully crafted evasion attacks at inference time, which can severely compromise model efficacy. Additionally, the high costs associated with training and prediction on massive graphs pose practical deployment challenges related to computational expenses. Furthermore, the poor interpretability of machine learning results, presents a critical operational hurdle. An alert from a “black box” model, lacking a clear explanation of the causal chain that triggered it, is difficult for security analysts to verify and act upon, undermining the goal of rapid, automated response. Addressing these concerns, this paper presents an extensive survey and organization of recent research on the application of learning technologies in provenance-based intrusion detection. It provides a systematic comparison of methods in terms of accuracy, efficiency, robustness, and result interpretability, which are crucial for system performance. This review summarizes the current state of research and highlights unresolved challenges, aiming to offer theoretical support and inspire future research and applications in this field. Ultimately, this work seeks to guide the community toward developing next-generation intrusion detection systems that are not only more intelligent and precise but also more scalable, resilient, and transparent.

Keywords intrusion detection; threat hunting; audit log analysis; provenance analysis; graph analysis; machine learning

1 引 言

网络空间中的攻击检测的本质是攻击方和防御方之间的智力博弈。从第一个计算机病毒被开发出来至今，攻防双方的技术均已得到多轮迭代升级。当前，面对无文件、离地（Live-off-the-Land，LotL）、混淆加密[1]、控制流劫持[2-3]等复杂隐蔽的攻击技术，传统的静态分析、敏感实体监控、网络侧流量分析等方法在大量场景中失去应有效果[4-5]。如图1所示，为了规避安全系统对Cron的监控，攻击者利用劫持正常系统控制流的技巧，掩盖其恶意行为。郭世泽院士等在“密态对抗”[6]一文中对问题进行了深入诠释，可归纳为网络空间中的隐身与反隐对抗问题。为对抗复杂的网络隐身技术，防御方需要对系统行为进行更全面的观测，以监控到恶意行为；需要有强大的分析能力，以及时准确地定位多样化的攻击行为，对攻击做出有效响应。

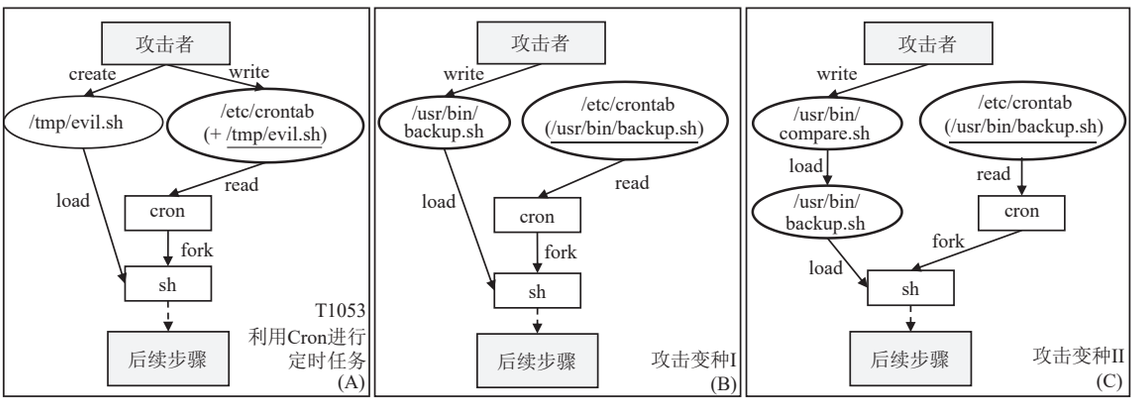

图1 溯源图表示的“T1053-利用Cron进行定时任务”攻击技术及其两个攻击变种((A)展示了一种攻击者利用Cron实现任务调度与持久化的简单方式;(B)和(C)展示了攻击者可通过劫持Cron任务的正常执行流构造攻击变种。在第一个变种中,攻击者重写了已存在于Cron任务列表的/usr/bin/backup. sh以注入攻击逻辑;在第二个变种中,攻击者进一步利用执行流,通过backup. sh触发的/usr/bin/compare. sh注入恶意代码。)

对网络攻击进行系统性观测是网络空间攻击检测的基础。观测可以从很多不同角度进行,以观测不同攻击特征,包括恶意文件的动静态分析[7]、流量分析[8]、日志分析[9]、以及多种形式的侧信道分析[10]等。然而随着攻击技术的演进,很多攻击特征被掩盖或不再体现,如恶意文件混淆遮蔽了恶意文件静态特征,而离地攻击①甚至不需要利用恶意文件就可以完成攻击。防御方亟需更新观测手段,对攻击中的不变性进行准确的观测与建模。2016年启动的“透明计算(Transparent Computing)②”项目,提出了“通过采集系统日志建模,使系统中发生的所有行为透明可观测”的研究思路,利用系统底层日志还原进程、文件等系统实体之间的信息流和交互关系,从而观测并建模系统行为。

上述过程采集的数据可以构建成图的形式,即溯源图(Provenance Graph)。该图包含了时间信息,可以反映攻击步骤之间的因果关系,因此也被称为因果图(Causality Graph)。溯源图能够准确表示进程等行为主体在系统层面的行为,如“打开文件”“建立远程链接”,并通过这些行为还原系统中的数据流和控制流,并将多阶段的攻击进行关联实现攻击行为建模,从而实现对复杂、隐蔽攻击的准确识别与检测。

然而,系统复杂的后台操作会引入海量的观测数据,对存储资源,尤其是高速存储提出很高要求。其中攻击行为相关的数据往往只占很小一部分,使得后续的入侵检测成为一个“大海捞针”式的任务。传统的非机器学习方法在一定程度上实现了数据压缩与高效搜索[11-13],但这些方法优化效果存在瓶颈且在不同场景下效果并不稳定。在后续的检测阶段,行为本身的复杂性又使区分攻击行为和正常行为变成一个复杂的建模问题。如何检测形态多变的攻击变种,并保持对针对性攻击的鲁棒性是一个开放的科学问题。为了保证效率,经典的方法通过图对齐[14-15]或图中基于标签传播的规则匹配[16-17]等方法实现大规模溯源图分析。虽然这些方法表现出了良好的性能和响应速度,但是其规则的表征能力往往比较局限,难以反映复杂攻击的多维特征,面对形态多变的攻击时表现出较差的鲁棒性。此外,为了保证检测到的攻击得到良好的处置,结果的可解释性与可响应性也是至关重要的,因此机器学习本身的黑盒特性会给检测带来额外的挑战。近年来,机器学习技术在推荐系统、流量预测等领域中解决冷启动、长程空间依赖性等问题[18-19],数据压缩、搜索[20]等任务的解决上有远超传统方法的表现。借鉴这些成果,越来越多的研究者开始将机器学习方法引入基于溯源分析的入侵检测领域[4,15,21-22],以提升入侵检测的各方面性能。如SHADEWATCHER工作引入推荐系统预测实体间交互偏好的思路,实现了低误报以及细粒度检测效果[23];STREAMSPOT工作则借鉴异构图相似性计算方法,实现了海量、高速数据流中高效的异常检测,系统的实时检测性能达到10万条事件每秒[24]。然而,使用智能算法并非简单地将溯源图套用到图神经网络或异构图模型中。溯源图本身海

① LOLBAS, https://lolbas-project. github. io/

② TC project, https://www. darpa. mil/program/transparent-computing

量实时的数据量、丰富的系统行为和语义信息以及入侵检测和响应任务的独特性都对智能化任务提出新的要求。由于解决问题的角度不同和提出攻击假设的差异，现有的智能溯源分析技术也各有侧重。

目前关于智能溯源分析的技术缺乏系统的整理、归纳、总结、分析。本文基于大规模的文献调研，对智能方法在入侵检测、对抗性攻击及审计数据处理等方面的应用情况进行了系统性分析，总结了对现有工作的洞察和仍然存在的挑战。综上所述，本文贡献点如下：

(1)本文是国内外首个智能溯源与入侵检测研究综述。近几年出现了许多将机器学习技术用于基于溯源的入侵检测的相关研究，对数据管理、入侵检测等检测的整体流程提出了大量新的想法，同时也引发了许多新的问题。作者在大量阅读相关文献，进行对比分析的基础上，结合相关研究与实践经验，第一次对此类工作进行了系统性的总结分析，提出了十余点洞察、挑战与展望(如表1所示)，希望为后续研究提供一些启示。

(2)基于溯源的入侵检测是被学术界和工业界均持续关注且寄予厚望的研究方向。然而学术界的论文往往做出一些理想化的假设使其难以满足业界实践的要求。本文基于作者在业界和学术界工作的经验，从实战角度出发，对已有研究工作从可处置性、响应效率等角度进行系统性对比分析，分析其在实战中可能面临的问题，希望为设计真正用于实战的系统提供新的视角。

(3)对抗性是网络安全的基本属性，攻击者必然会针对检测系统设计方案展开攻击。机器学习技术的使用在给检测带来了新能力的同时，也引入了新的攻击面。针对机器学习的数据投毒、模型逃逸等攻击也给传统的入侵检测领域带来困扰。本文从对抗的角度出发，总结多种对抗性攻击的方法，对已有的系统进行鲁棒性的评估，希望为检测系统的设计提供新角度和理论支持。

表1　洞察、挑战与展望

编号	洞察/挑战/展望	章节
洞察1	基于溯源图的入侵检测需要从数据采集、压缩、管理到检测、调查、响应的复杂系统支撑。智能算法在各个阶段都有广泛的应用，但在算法设计时应综合考虑对系统的影响，实现系统全局效果的最优化	3.1
洞察2	基于序列的检测方法计算开销较低，能够在有限资源条件下实现高效检测。然而由于其难以分析更广泛的关联关系，检测效果(准确率、鲁棒性、结果可解释性等)相应较差	4.1
洞察3	基于节点/边的检测方法能够提供细粒度告警信息，缓解审计人员过滤无关信息的开销	4.2
洞察4	基于路径的检测方法能够关联到长时期系统实体间的依赖关系，具有更好的准确率和鲁棒性。同时，此类方法天然契合溯源分析中经典的基于标签传播的计算框架，能够实现更快的响应和更低的计算开销	4.4
洞察5	入侵检测系统中的设计应尽可能保持简洁，额外的模块将向攻击者暴露更多的攻击面，影响整体系统面对对抗性攻击时的鲁棒性。尤其是数据过滤等可能会移除数据的模块的引入应保持谨慎	5.1
洞察6	任何防御手段都存在绕过风险，基于溯源的检测技术研究可以有效提升绕过难度，提升系统安全性	5.3
洞察7	自然语言处理技术能高效提取非结构化的攻击报告中的知识，并结构化表示，有效提升威胁分析效率	6.3
挑战1	溯源数据规模大，但涉及的攻击数据少，存在数据不平衡的问题，加上攻击本身的多样性问题，给准确高效的攻击或异常行为建模带来挑战	3.2
挑战2	溯源数据集没有统一的标准，涉及攻击和背景数据质量参差不齐，给有效的系统评估带来挑战	3.2
挑战3	尽管 P-EDR 已经得到了广泛的业界应用，“有限计算资源 vs. 复杂计算任务”、“受限适应能力 vs. 多样攻击场景”、“结果难以解释 vs. 告警疲劳问题”和“有限规则集 vs. 未知攻击技术”等问题仍然很大程度地限制了其效果	3.4
挑战4	攻击图可以高效、准确地描述入侵流程。但如何从大规模溯源图中准确划分出攻击图是一个挑战	4.3
挑战5	机器学习算法在为溯源分析引入更强的能力的同时也引入了额外的风险。如何有针对性地优化机器学习算法，提升其在入侵检测时的鲁棒性是一个挑战	5.2
展望1	机器学习模型寻找和表达特征的能力，为准确、高效的表征攻击模式提取，提升溯源检测的准确率；以及高效的数据压缩和搜索实现，有效提高分析效率及降低分析开销，提供了新的思路和解决方案	3.3
展望2	溯源图中存在大量冗余、无关信息，通过合理利用机器学习算法可以实现对关键信息的高效的筛选、总结与压缩，有效提升溯源数据存储、查询和处理效率	6.2
展望3	入侵检测是一个系统性问题。当前基于智能溯源分析的入侵检测各模块在部分属性上实现了比较理想的效果，但这些模块之间往往不能简单地组合到一起。如何合理组合使用这些模块，达到整体效果最优是一个开放性问题	7.2

2 综述方法

2.1 文献收集与分析

作者此前的工作[25]、伊利诺伊大学香槟分校Inam等人的知识总结工作[21]以及Zipperle等的综述[26]对溯源分析在入侵检测和威胁分析领域的理论和问题进行了总结和讨论，梳理了溯源分析从日志采集，到数据的压缩与存储，再到溯源建模与入侵检测分析的全流程中各阶段的挑战与传统解决方案。近年来，大量人工智能方法被引入溯源分析领域，为该领域提供了新的研究机会和思路。

本文特别关注那些将机器学习技术与溯源分析相结合的创新性工作，这些研究不仅开创了溯源安全分析的新范式，也为传统溯源分析入侵检测提供了新的技术思路。为确保研究的规范性和可重复性，本文在文献收集过程中采用以下纳入标准：(1)文献提出工作基于溯源方法，且研究成果对安全分析人员具有实际指导意义。(2)文献涉及机器学习方法在溯源图分析中的应用，包括对溯源图数据的建模、特征学习或攻击检测。(3)研究工作应在方法创新性或应用效果上具有显著贡献。

同时，本文采用以下排除标准：(1)文献仅关注系统实现或工具开发，未涉及数据溯源相关技术。这类文献虽然可能涉及安全分析，但缺乏对溯源数据和溯源图的深入研究，与本文的研究重点存在显著偏差。(2)文献仅应用传统统计方法或规则匹配技术，未涉及机器学习技术。这类方法在一些场景下存在其优势，但未能充分利用机器学习技术的优势，难以应对复杂的攻击场景分析需求，并非本文研究重点。

根据上述标准，本文通过三个步骤筛选文献：(1)根据研究方向确定检索的关键词如表2所示，检索的数据源包括IEEE Xplore Digital Library、SpringerLink Online Library、ACM Digital Library等。检索时间区间定义在2016年至2024年。(2)按照标题、关键词、摘要、结论和来源的优先顺序对上述文献进行筛选，其标准为：①根据发表文章的单位、团队，录用文章的会议、期刊，以及引用量等信息综合筛选网络空间安全、系统和软件工程领域高水平论文；②与智能化溯源分析及入侵检测技术有关。(3)由第一作者和第二作者对上述文件进行全文排查，并对这些文献利用Google Scholar和Web of Science的引文索引功能进行前向后向追踪(Snowballing)的方式再进行补充和筛选。

表2 检索关键词及逻辑

关键词	溯源 + (学习|智能|嵌入|表示|知识) + (攻击|威胁|入侵|恶意)+(检测|取证|调查|狩猎)
Keywords	Provenance + (Learning|Intelligence|Embedding |Representation|Knowledge) + (Attack|Threat|Intrusion|Malicious) + (Detection|Forensic|Investigation|Hunting)

通过上述步骤对文献进行筛选后，得到一百余篇相关与智能化溯源分析用于入侵检测相关的文献。图2展示了本文总结的入侵检测、威胁情报分析及其他溯源分析领域智能化相关文献发表年份和研究问题分布情况。可以看出，该研究方向受到了学术界的广泛关注，且处于上升趋势。尤其是，基于溯源图的智能入侵检测算法设计得到了越来越多的研究，是一个开放且活跃的研究问题。需要说明的是，尽管我们采用了系统化的检索方法，但该领域研究十分活跃，新的研究工作和业界应用案例层出不穷，任何文献综述都不能规避遗漏风险。

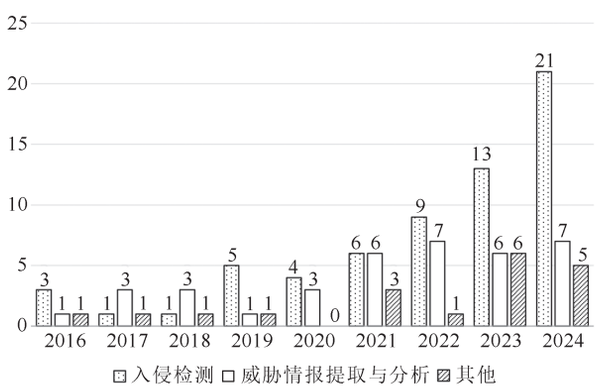

图2 论文分布统计

2.2 研究思路和框架

如图3所示，基于溯源图的入侵检测分析需要构建从数据采集到数据的压缩与存储，再到检测响应的复杂系统，涉及日志可信、流计算、语义分析等一系列核心技术，是一个复杂的系统性研究工作。本文围绕该系统架构，对智能溯源分析全周期进行系统性的综述分析，总结研究和应用过程中的方法、效果和经验。首先介绍溯源分析在入侵检测领域的应用相关背景知识(第3节)，然后围绕以下研究问题展开讨论：

研究问题1：(第3节)基于溯源分析的入侵检测方法的业界渗透率如何？当前主要面临哪些挑战？

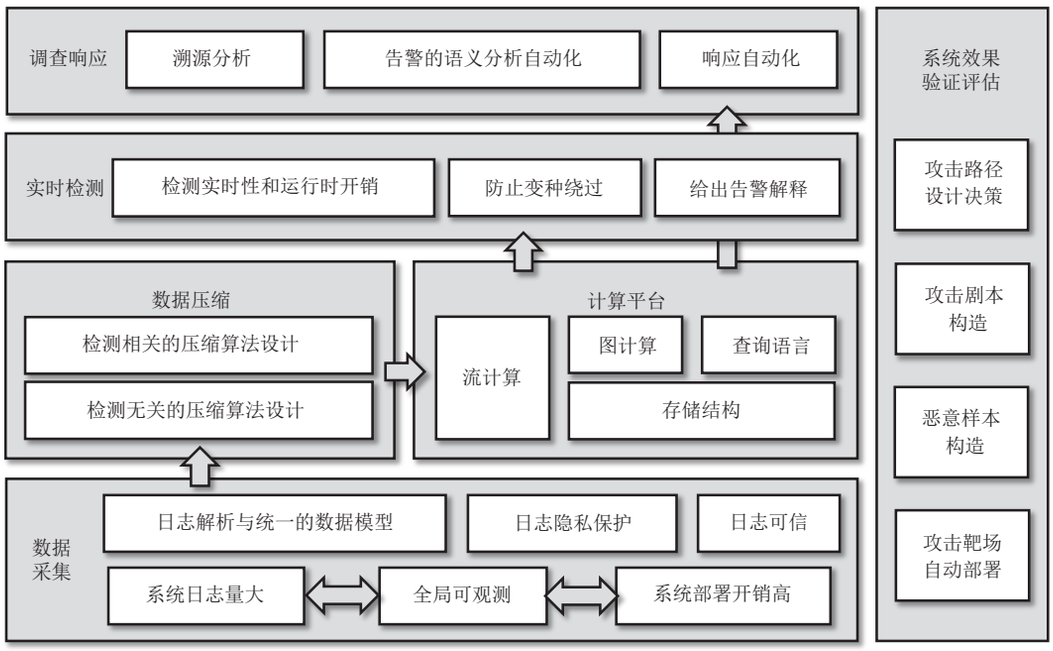

图3　基于溯源图的入侵检测与威胁分析技术研究体系

智能溯源分析方法将如何解决这些问题?

研究问题2:(第4节)基于机器学习的溯源图中攻击检测方法如何对溯源图进行建模和分析? 这些方法能否有效区分溯源图中的攻击行为和正常行为?

研究问题3:(第4节)述检测算法的性能开销以及结果的可解释性如何? 能否实现在线的检测,支撑及时有效的威胁响应与处置?

研究问题4:(第5节)针对基于机器学习的溯源图中攻击检测,攻击者有哪些对抗性的攻击策略? 现有检测方案对这些策略的鲁棒性如何?

研究问题5:(第6节)智能算法如何提升溯源图的数据压缩和查询效率? 这些效率提升是否会带来准确率等其他方面的性能损失?

研究问题6:(第6节)智能化方法还在哪些其他方面帮助提升基于溯源分析的入侵检测效果?

最后对综述内容进行总结与讨论,归纳当前的仍困扰学术界和业界的开放性研究问题,提出潜在的研究方向(第7节)。

3　基于溯源图的入侵检测系统及其智能化

3.1　基于溯源图的入侵检测相关基本概念

随着网络攻击的复杂性和数量的持续增加,安全审计人员仅依靠组织内部现有的检测工具,已难以在海量系统日志中高效捕获攻击行为,这对内部安全团队构成了严峻挑战。为应对此问题,学术界引入了溯源图的概念[27-28]。如图4所示,溯源图可以将日志数据构建为因果关联的操作关系图,直观清晰地描述网络攻击或其他行为的路径,使审计人员能够快速定位攻击的根本原因或评估攻击的影响范围。溯源分析系统通常利用系统监控工具(如EBPF①和ETW②等)从内核层面收集系统调用生成的审计事件,并基于这些事件构建溯源图,以直观展示系统实体之间的交互活动。广义的溯源图是一种有向图,其中节点代表系统实体(例如进程、文件和IP地址),边则表示这些实体之间的控制流和数据流关系。溯源图可以表示为一组事件四元组〈起点(s),终点(d),操作(o),时间(t)〉的集合。表3列出了常用的事件类型,这些事件作为边首尾相连就构造了动态扩展的溯源图。

如图3所示,基于溯源图的入侵检测、威胁狩猎等是一个复杂的系统性工程,并逐渐发展成包括多个相互支撑的技术模块的复杂系统平台,如数据采集、数据压缩、数据计算平台、实时检测与调查响应、威胁情报的提取与分析及系统验证评估等。而机器学习技术在其中多个模块都表现出很好的利用效果和利用前景。但在算法设计时应综合考虑对系统的影响,实现系统效果的最优化,避免单一模块优化而

① eBPF. https://ebpf.io/
② Event tracing for windows. https://docs.microsoft.com/en-us/windows/win32/etw/about-event-tracing.

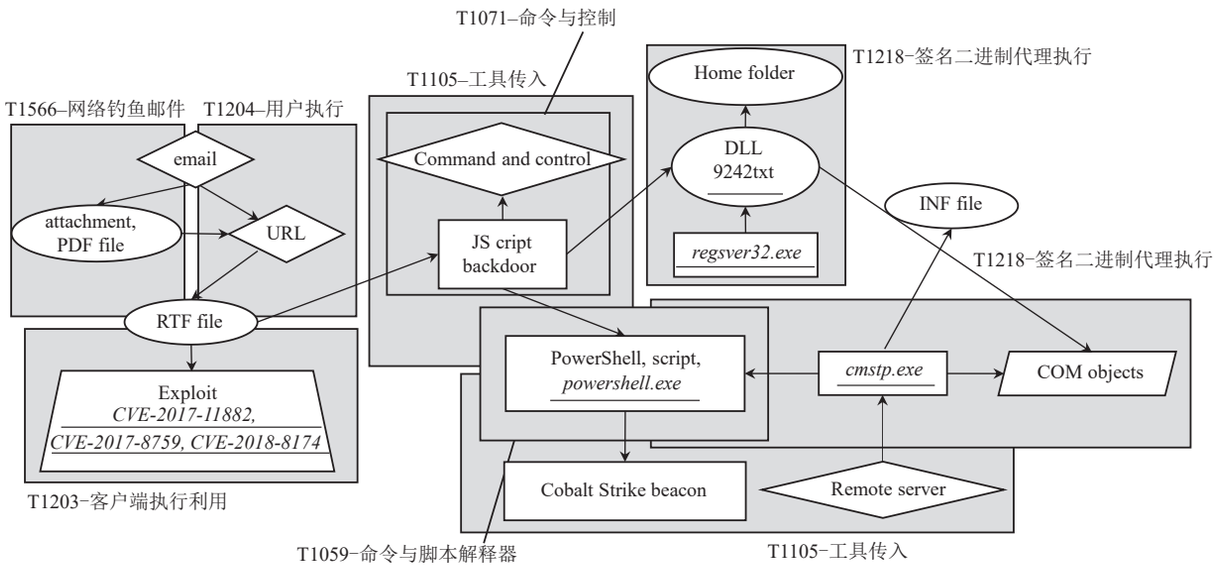

图4 溯源图表示的Cobalt攻击流程及涉及的攻击技术标注(Cobalt攻击利用钓鱼邮件,诱导用户执行恶意附件或链接,利用Office漏洞(如CVE-2017-11882)触发代码执行;随后通过JScript后门或DLL文件投递恶意工具,并滥用签名进程进行执行;最终部署Cobalt Strike Beacon建立远程C2通信,利用脚本解释器执行命令、横向移动及持久化,并将完整攻击过程映射至MITRE ATT&CK技术。)

表3 溯源图中事件的定义

起点(s)	终点(d)	操作(o)	图例
进程	文件	写/创建文件	进程→文件
文件	进程	读/加载文件	文件→进程
进程	进程	启动/终止进程	进程→进程
进程	网络	发送消息	进程→网络
网络	进程	接收消息	网络→进程

注:在本文绘制图例中,进程节点用方框表示,文件节点用椭圆表示,网络节点用菱形表示,漏洞利用用梯形表示,其他实体用灰底的方框表示。

影响其他模块效果,如数据压缩可能丢失关键信息,影响后续检测溯源分析准确率。

洞察1:基于溯源图的入侵检测需要从数据采集、压缩、管理到检测、调查、响应的复杂系统支撑。智能算法在各个阶段都有广泛的应用,但在算法设计时应综合考虑对系统的影响,实现系统全局效果的最优化。

3.2 开源数据采集工具和数据集

机器学习算法和模型的效果很大程度上依赖于训练数据的质量和规模。网络攻击相关的数据集较图像处理等主流人工智能研究领域的数据集更难以收集与标注,而溯源日志本身采集的难度和数据量进一步提升了数据收集的难度。我们总结了常用的开源溯源数据采集工具和数据集,帮助了解该研究领域及开展后续研究。

表4列举了系统底层审计工具及典型数据集的核心信息。审计工具中,SPADE作为分布式日志审计工具,支持Windows ETW, Linux Syslog等多种数据接入,并提供统一的溯源图解析、生成和处理接口。CamFlow则是基于Linux安全模块(LSM)的捕获工具,可灵活适配不同应用场景。

在数据集方面,DARPA TC E3、E5数据集分别来自DARPA第三次和第五次红蓝对抗演习(Engagement)。演习主要针对企业关键服务(如Web服务器、电子邮件服务器和SSH服务器等)和个人环境(包括个人电脑、路由器等)发起了从初级的文件投放和未加密C2通信逐步升级到更复杂的技术,如内存反射加载、系统调用规避和模块化攻击工具链构成的一系列模拟攻击。该演习希望通过实战化的演练验证数据项目团队数据采集和检测算法的能力,在项目执行期内共组织了5次,其中第三次和第五次的数据集被公开了,得到了学术界的广泛关注和利用。例如,在入侵检测领域,FLASH[29]和MAGIC[30]等研究利用这些数据集评估检测的准确率和效率;在溯源取证调查方面,NODOZE[31]和DEPIMPACT[32]等研究则基于这些数据进行了深入的攻击行为分析和影响评估。

DARPA OPTC数据集记录了7天内500台Windows主机的活动,其中前4天为良性活动,后3天则混合了良性和高级持续性威胁(APT)攻击行为。在这三天的攻击活动中,每一天都展示了不同

表4　基于溯源图的入侵检测开源数据采集工具和数据集

	数据集/工具名	平台	持续时间	来源	数据内容
采集工具	SPADE[1]	Windows Linux MacOS	/	哥伦比亚大学	基于Event Tracing for Windows，Linux Auditd，OpenBSM的跨平台数据采集工具，支持Linux、Windows、macOS等多个平台，对系统调用、文件操作、网络活动等细粒度日志进行采集
	CamFlow[2]	Linux	/	剑桥大学	基于Linux Security Modules实现的Linux内核的实时数据溯源框架，支持进程、文件和网络活动的实时数据采集
数据集	DARPA TC E3[3]	Windows Linux FreeBSD	14天	DARPA	E3数据来自DARPA TC项目组织的第三次红蓝团队练习。期间，红队针对企业关键服务，包括WEB服务器、电子邮件服务器等，发起了包括水坑攻击、数据窃取等一系列攻击
	DARPA TC E5[4]	Windows Linux FreeBSD	9天	DARPA	E5数据来自DARPA TC项目组织的第五次红蓝团队练习，涉及初级攻击技术，包括文件投放等，到高级攻击技术，包括内存反射加载、系统调用规避等
	DARPA OpTC[5]	Windows	3天	DARPA	OPTC数据集涵盖了7天内500台Windows主机的良性活动，以及3天混合了良性和APT活动的数据，使用了Powershell Empire等开源攻击实施攻击
	PKU ASAL[6]	Windows Linux	2天	北京大学	ASAL数据由北大应用安全实验室构建，模拟了APT29、Side Wide等组织的攻击行为，涵盖恶意软件投放、权限提升等技术
	PASS[7]	Linux	/	清华大学	PASS则由Start Bootstrap构建的APT攻击链生成，包含Vsftp(CVE-2011-2523)、Samba(CVE-2007-2447)等典型漏洞利用

注：[1] SPADE：https://github.com/ashish-gehani/spade [2] CamFlow：https://camflow.org/ [3] DARPA TC：https://github.com/darpa-i2o/Transparent-Computing/blob/master/README-E3.md [4] DARPA TC：https://github.com/darpa-i2o/Transparent-Computing [5] OpTC：https://github.com/FiveDirections/OpTC-data [6] PKU ASAL：https://github.com/PKU-ASAL/Simulated-Data [7] PASS：https://github.com/yulaixie/yulaixie.github.io/tree/master/datasets/pass

类型的攻击技术。例如，第一天模拟了Powershell Empire的攻击场景，包括初始入侵、横向移动和权限提升等阶段。PKU-ASAL数据集通过三台主机复现APT29等高级威胁中独立攻击步骤生成，因此攻击链并不完整，在NODLINK[33]工作中，它被用于评估细粒度检测效果。PASS则利用Metasploit工具链构建多阶段攻击链(如漏洞利用、权限绕过)，完整覆盖入侵、横向渗透到数据泄露流程。

当前公开数据集广泛覆盖了多阶段攻击场景(如APT渗透、漏洞利用、数据窃取等)，为评估检测系统的跨场景泛化能力提供了基础。然而，不同团队采用的采集工具存在差异，收集到的日志事件类型、数据维度与粒度并非完全一致(如进程行为、网络通信、权限变更等字段的缺失或冗余)，在利用时需要额外的处理，且存在非标准化处理的问题。此外，数据集中涉及的攻击路径也存在很大的差异，特别是部分数据集中攻击行为并不能组成完整因果关联的攻击链条，给检测模型的迁移能力验证与横向对比带来了挑战[34-35]。另外，这些数据集往往也缺乏明确、可复现的Ground Truth标注流程，如DARPA TC系列仅提供攻击场景描述，而未与具体日志条目对应，导致测试者往往需要自己定义标签，造成了验证结果产生主观差异，以及不同团队检测系统表现出性能的不一致的问题[36]。当前，业界急需一个规模更大，覆盖攻击更多，数据质量更高，标注更为清晰的开源数据集。

挑战1：溯源数据规模大，但涉及的攻击数据少，存在数据不平衡的问题，加上攻击本身的多样性问题，给准确高效的攻击或异常行为建模带来挑战。

挑战2：溯源数据集没有统一的标准，涉及攻击和背景数据质量参差不齐，给有效的系统评估带来挑战。

3.3　传统方法的局限性

传统的溯源检测方法通常依赖于固定策略(Policy)来识别恶意行为。这些方法通过基于已知攻击行为知识和异常行为模式定义事件匹配规则，并用于检测潜在攻击。然而，由于网络攻击的复杂性和多样性，各种攻击变体经常超出规则集的覆盖范导致传统溯源检测系统在应对变体攻击或未知攻击时显得力不从心。此外，传统的数据压缩算法有明显的压缩倍率瓶颈，难以应对大规模系统中复杂行为带来的海量观测数据。近年来，机器学习技术的发展给基于溯源分析的入侵检测领域带来了新的机遇。一是机器学习模型强大的特征表达能力有可

能更加准确、高效地表征攻击模式，区分正常行为与攻击行为，提升溯源检测的准确率。二是机器学习模型能够找到海量数据里的一般特征，从而实现更高效的数据压缩和搜索，有效提高分析效率，降低分析开销。具体来说，智能算法可能在以下一些方面得到具体应用：

(1)检测与调查响应。检测和调查响应是入侵检测的基本任务。传统的检测方法往往通过预定义的攻击或异常模式进行威胁行为识别，包括异常的数据流出、典型的攻击持久化行为等。然而，传统模式的泛化能力和鲁棒性较差，难以发现以零日攻击为代表的未知攻击模式。基于机器学习的检测方法能够用更加一般和鲁棒的形式表示复杂溯源图中的行为特征，并将其转换为不同粒度的低维向量表示。随后，通过利用历史日志数据训练的基线模型或攻击行为模型，进行威胁行为的检测与识别。此类方法对未知攻击和变体攻击较强的适应能力，有效弥补传统方法的不足。

(2)数据存储与查询。合理的数据管理，包括数据的存储结构、压缩方式、查询接口设计能有效提升后续分析阶段的效率。受限于溯源图庞大的规模，传统的溯源图存储方法常通过删除冗余的因果事件等手段减少内存占用。然而，这些方法本质上仍然未改变图数据的存储格式，效果比较受限。基于机器学习的压缩方法则采用更低维度的向量替代字符信息，并利用机器学习模型替代传统的图数据库索引。相比于传统的压缩方法，机器学习模型显著降低了存储溯源图的内存需求，并提升了图查询的效率。

(3)威胁情报提取与告警解释。为更好地解释攻击事件，安全从业者以网络威胁情报(Cyber Threat Intelligence)报告的形式积极总结和交流组织之间有关攻击的知识。由于威胁报告以自然语言文本编写，传统的方式需要繁琐的手动网络威胁情报恢复工作。而机器学习技术可以自动化地提取报告中的知识并以标准化的、机器可读的形式表示，形式包括威胁指标IoC、攻击知识图等。自动化的提取方式有效提升威胁知识提取、分享的效率，帮助更好地解释实际遇到的多样化的网络攻击。

展望1：机器学习模型寻找和表达特征的能力，为准确、高效的表征攻击模式提取，提升溯源检测的准确率，以及高效的数据压缩和搜索实现，有效提高分析效率及降低分析开销，提供了新的思路和解决方案。

3.4 业界的应用与反馈

一项涵盖多家安全公司访谈和调查的研究表明，业界已经充分认识到溯源图在入侵检测中的优势[37]。经验丰富的审计人员能够轻松理解溯源数据，即使这些数据仅包含底层的系统审计事件。与传统的终端检测响应系统(Endpoint Detection and Response，EDR)相比，P-EDR(Provenance-Based EDR)系统的优势在于它能通过溯源分析自带的因果关系(Causality)，在检测和后续的调查取证步骤中能建立系统行为和实体构成的溯源图[16,31]，从而对全局攻击行为进行关联分析，实现更加准确的攻击行为检测，同时提供更丰富的攻击语义。

当前，微软、IBM、华为等国内外头部科技企业正加速将攻击溯源能力整合至EDR及安全响应中心(Security Operation Center，SOC)体系中。这些成果表明，P-EDR系统能有效提升传统的EDR系统能力，成为应对高级网络攻击的核心防御工具。特别是在有效解决无文件攻击和离地攻击等复杂场景时，P-EDR展现了巨大的优势。在渗透测试等对抗性测试中，P-EDR已经成为了入侵者面临的巨大挑战，从侧面说明了其防护有效性。MITRE等组织围绕溯源分析提出了ATT&CK攻击矩阵(Matrix)①、攻击流模(Attack Flow)②等开源框架，提供了标准化的网络攻击行为描述和建模方法，可将网络空间中的攻击行动拆解为原子化的战术技术，并按实际攻击场景串联为行为序列，并通过关联系统日志实现攻击路径回溯。

表5列出了已集成溯源检测功能主流安全厂商的安全产品，并适配本地化、云原生等多样化场景。然而，为保护知识产权，多数产品未公开详细的系统组成。就公开可获取的资料来看，华为HiSec EDR通过进程树分析实现系统级进程实体溯源，但其事件跟踪类型较为局限。青藤猎鹰平台的核心架构包含数据收集、数据分析(集成MITRE ATT&CK框架的查询解析引擎及机器学习行为分析引擎)和Web控制台，检测逻辑部分依赖外部输入。其他产品如深信服可扩展检测响应系统(Extendable Detection and Response，XDR)等系统组成也主要是由规则引擎和机器学习引擎结合进行检测。尽管这些产品在检测性能上表现突出，但它们的早期过滤

① ATT&CK Matrix，https://attack.mitre.org/matrices/enterprise/

② Attack Flow，https://ctid.mitre.org/projects/attack-flow

表5　部分应用溯源分析的安全产品

安全厂商	产品	溯源功能	主要面向场景
IBM	QRadar	支持进程级攻击路径溯源和威胁根因分析	混合云环境等
CrowdStrike	Falcon Endpoint Security	支持攻击时间线与攻击链溯源	混合云环境等
SentinelOne	Singularity Platform	提供主机内根因分析与攻击路径溯源	混合云、边缘计算等
Microsoft	Defender for Endpoint	支持追溯攻击路径并提供修复建议	Azure生态等
华为	HiSec Endpoint	支持进程级攻击路径溯源和威胁根因分析	云防护、本地安全等
青藤	青藤猎鹰	支持用户行为与资产监控的攻击回溯	云防护、容器安全等
深信服	可扩展响应平台 XDR	支持追踪系统内部威胁和异常操作链	端网协同环境等

机制通常仅关注特定事件以降低开销。这导致常规事件可能被忽略，从而被攻击者利用形成检测盲区，成为这些系统的通用缺陷。除此之外，产业界实际应用过程中还存在更直接和严峻的挑战，包括：

(1)有限计算资源 vs. 复杂计算任务。大多数安全团队认为P-EDR系统的高运营成本是其推广应用的主要障碍。这些成本包括客户端和服务器端的计算资源消耗，以及警报分类和攻击调查所需的人工投入。虽然学术界通常将检测精度视为最重要的指标，但实际操作中，大多数安全团队更关注系统的经济性和可行性。尽管P-EDR系统通常具备较高的检测精度，并显著降低了误报率，但其昂贵的运营成本使得许多安全团队望而却步。

(2)受限适应能力 vs. 多样攻击场景。P-EDR系统在现实世界的应用场景比实验室环境更加复杂，它们可能会面对多种未知的攻击行为以及此前未观察到的良性行为，从而显著增加溯源检测的难度。尽管已有研究尝试通过跨主机行为的收集来扩大基线学习的覆盖范围[38]，但这些方法仍然依赖于相对简单的泛化机制进行匹配。这种局限性导致P-EDR系统在处理缺乏显著特征的实体时容易产生误报，进一步降低了检测过程的准确性与可靠性。

(3)有限规则集 vs. 未知攻击技术。在安全研究和攻防对抗中，如何检测未知攻击(如攻击变体或0日威胁等)一直是一个难题。一种常见的思路是采用基于异常检测的方法，利用机器学习模型等方式学习历史日志数据的特征，并在发现与所学特征明显不符行为时触发告警。然而，这种方法存在误报率高、可解释性差等局限性，在面对海量图数据时，分析的开销较高，效果较差。另一个思路是对基于规则的检测系统进行改进，利用溯源图较好的攻击语义表达能力，通过泛化的检测规则等提升对攻击变种的覆盖能力，并保留查询接口支持后续查询规则的扩展。

(4)结果难以解释 vs. 告警疲劳问题。告警疲劳问题是业界面临的重要挑战。当前，检测系统给出的告警需要人工的确认与响应处置。海量且难以解释的告警导致分析人员疲于应对，而无法及时处置的告警造成的破坏与检测系统的漏检无异。

挑战3：尽管P-EDR已经得到了广泛的业界应用，“有限计算资源 vs. 复杂计算任务”、“受限适应能力 vs. 多样攻击场景”、“结果难以解 vs. 告警疲劳问题”和“有限规则集 vs. 未知攻击技术”等问题仍然很大程度地限制了其效果。

4　基于机器学习的入侵检测算法设计

溯源图是一种利用审计日志来描述系统行为的有向图，其中节点代表系统中的实体(如进程、文件、网络等)，而不同的系统调用或操作则通过有向边连接，边上的时间戳刻画了溯源图随时间演变的过程。受到子图异质性[39]的启发，在主机上利用溯源图实现入侵检测可以转化成溯源图建模过程的异质性问题。基于学习的溯源检测旨在发现那些不常见或不典型的行为模式，这些异常行为通常通过其与图主要部分的结构差异来表现，表7系统性地总结了主要工作的编码、预测及工作模式。给定溯源图 $G=(V,E)$，其中 $V=\{v_j\}_{j=1}^{|V|}$ 表示节点集合，$E=\{e_j\}_{j=1}^{|E|}$ 表示边集合。

根据建模方式不同对机器学习在入侵检测中的应用进行区分：(1)基于日志序列的入侵检测通过将图 G 按时间演变划分为时间连续的序列 G_{seq}，并利用机器学习方法构建嵌入函数 $\eta_{\text{seq}}=G_{\text{seq}}\to\mathbb{R}^d$，将每个序列映射成 d 维度的向量 $\boldsymbol{z}\in\mathbb{R}^d$。(2)基于节点特征的入侵检测通过识别与历史行为偏差较大的节点，使用嵌入函数 $\eta_v=v\to\mathbb{R}^d$，将每个节点 $v\in V$，及其邻域分布映射为 d 维度的向量 $\boldsymbol{z}\in\mathbb{R}^d$。(3)基于路径特征的入侵检测将图 G 分割为一组路径子图 G_p，通过嵌入函数 $\eta_p=G_p\to\mathbb{R}^d$，将每个子图映射为

d维度的向量$\boldsymbol{z}\in\mathbb{R}^d$。得到不同建模方式生成的向量信息后，进一步根据日志过往的正常历史行为，学习其中子图预测函数$\varphi(\boldsymbol{z})$作为基于向量的度量来确认是否出现了异质的情况。为了解决这个问题并实现更大的通用性，使用一个统一的指标，定义如下：

$$Z^*=\{\cup\boldsymbol{z}\mid\exists\boldsymbol{z}\in\mathbb{R}^l,\varphi(\boldsymbol{z})>\alpha\} \tag{1}$$

其中，α表示告警的阈值，Z^*表示所有满足异常条件的向量集合。图5(a)至图5(d)分别对应四种不同的检测分类，日志序列特征、节点/事件特征、路径特征和子图特征学习。表6系统性地总结了不同种类检测的优势、劣势及应用场景。在4.1节至4.4节详细介绍了它们更具体的研究工作。

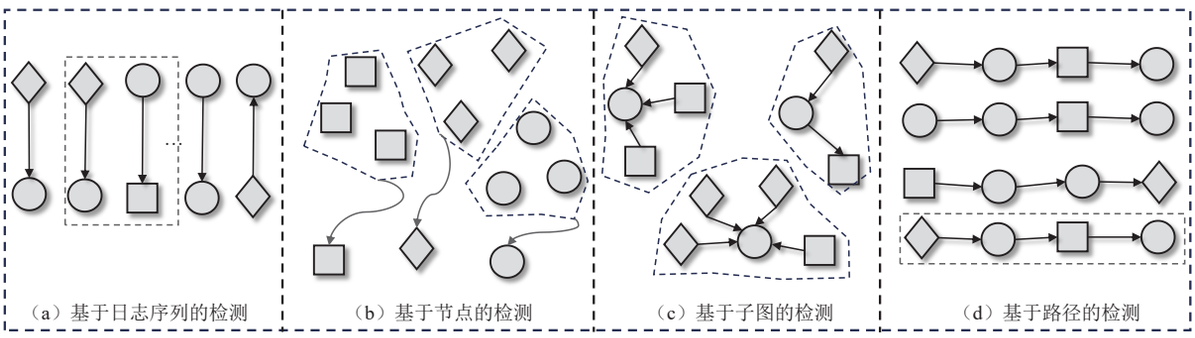

图5 不同种类的溯源检测方法

表6 不同种类的基于溯源分析的入侵检测系统对比

检测思路	优 势	劣 势	应用场景
基于序列	分析计算开销低，适合在资源有限环境部署	难以分析全局行为关联，难以检测长时空攻击，且可解释性不足	适用于时序关联的攻击分析，无需复杂依赖计算的场景
基于节点	分析开销低，且提供单个实体的细粒度告警，降低人工分析过滤成本	准确率较低，且告警可解释性不足，无法支持溯源和响应建议	适用于快速生成细粒度告警，无需上下文关联分析的简单攻击场景
基于路径	关联攻击行为的长期依赖关系，检测准确性高	大规模路径分析面临路径爆炸问题，分析复杂度较高	适用于长周期多阶段攻击检测，需要全链路因果分析场景
基于子图	高效、准确地描述入侵流程，可解释性较高	易受良性数据干扰，子图分割会导致检测精度损失	适用于处理结构化攻击，表征多实体关联的攻击的场景

4.1 基于日志序列学习的入侵检测

基于日志序列学习的入侵检测方法关注系统日志中事件的时序性和连续性，通过分析日志条目之间的顺序、时间、上下文依赖等信息对日志进行建模来识别潜在的恶意行为。

DEEPLOG[40]是基于日志序列进行入侵检测的经典工作。该工作利用长短期记忆网络(Long Short-Term Memory，LSTM)对正常日志序列的模式进行建模，从而检测潜在异常行为。然而，该工作仅关注时序问题而忽视了日志条目间的其他重要关联，如用户日常行为模式间的逻辑关联。为克服这一局限，中国科学院刘福承等人提出了一种基于异构图嵌入的创新性检测框架LOG2VEC[41]。该框架通过自定义的异构图集成日志条目之间的多维关联关系，并通过图嵌入算法学习日志的向量表示，最后设计定制的模型将相似的恶意操作聚类。该方法显著提升了检测系统对复杂攻击场景的适应能力以及检测的准确性。然而，其在处理长期依赖关系时会遇到依赖爆炸问题，导致计算复杂度显著增加。

在上述工作的基础上，ATLAS[42]，通过消除冗余的节点和边提取关键攻击和非攻击序列，并利用LSTM学习序列模式以识别高级持续性威胁攻击的关键步骤。该方法在保留时间信息的同时，通过因果分析(Causality Analysis)对攻击相关数据进行了筛选，减少了每次分析的数据量，在降低了开销的同时提升检测准确率。整体而言，日志的时序信息为入侵检测提供了重要信息，但关联窗口有限，也无法有效对事件间其他维度特征进行有效关联，难以捕捉长时空攻击的全局信息。相对的，后续的工作更加重视基于因果关系给出的攻击语义，从节点/事件、路径、图等粒度对溯源图进行分析。

洞察2：基于序列的检测方法计算开销较低，能够在有限资源条件下实现高效检测。然而由于其难以分析更广泛的关联关系，检测效果(准确率、鲁棒性、结果可解释性等)相应较差。

表7　智能溯源分析和检测系统对比

类别	系统	时间	实时/离线	编码方式		预测方式	
				编码特征	编码模型	预测算法	预测模型
基于序列	DEEPLOG[40]	2017	实时	序列	LSTM	分类	自定义规则
	LOG2VEC[41]	2019	离线	节点	Word2Vec	聚类	自定义规则
	ATALAS[42]	2021	离线	节点 边 时间	Word2Vec	序列回归	LSTM
基于节点	SIGL[43]	2021	离线	节点 邻域	GraphLstm	分类	Jenks
	SHADEWATCHER[23]	2022	离线	节点 边 邻域	TransR GAT	内积预测	自定义规则
	THREATRACE[44]	2022	实时	节点 邻域	GraphSage	分类	Multi-model
	PROGRAPHER[45]	2023	实时	节点 邻域	Graph2Vec	分类	TextRCNN
	PROVCTDG[46]	2023	实时	节点 时间	TGN	时间链接预测	自定义规则
	NODLINK[33]	2024	实时	节点	Fasttext	回归	VAE
	MAGIC [30]	2024	离线	节点 邻域	GAT	相似度计算	KNN
	FLASH [29]	2024	实时	节点 边 邻域	Word2Vec GNN	节点分类	XGboost
	R-CAID[47]	2024	离线	节点 邻域	Doc2Vec GNN	节点聚类	K-means
	ORTHRUS[48]	2025	实时	节点 邻域	OneHot GNN	节点聚类	K-means
基于路径	PROVDETECTOR[49]	2020	离线	节点 边	Doc2Vec	局部异常离群	LOF
	PG-AID[50]	2024	离线	节点 边	Doc2Vec	自回归	transformer
基于子图	STREAMSPOT[24]	2016	实时	节点 邻域	SimHash	聚类	K-medoids
	UNICORN[51]	2020	实时	节点 边 邻域	LSH	聚类	K-medoids
	APT-KGL[52]	2022	实时	节点 邻域	HGAT	分类	R-GCN
	PROVG-SEARCHER[53]	2023	离线	节点 邻域	GNN	图匹配	自定义规则
	KAIROS[54]	2024	实时	节点 邻域	HFH TGN	回归	MLP
	MEGR-APT[55]	2024	离线	节点 邻域	GNN	图匹配	GED
	TREC[56]	2024	离线	节点 邻域	HAN	距离度量	Siamese

4.2　基于节点/事件特征学习的入侵检测

在以对抗网络攻击为目标的安全事件响应流程中，入侵检测仅是第一步，粗粒度的特征表示和检测结果难以提供充分的攻击语义，使得后续针对性的应急响应变得更为困难，因此基于节点和事件等细粒度特征入侵检测系统受到了研究者的关注。Han等人提出的SIGL[43]是最早使用图神经网络对节点信息进行表示的检测系统之一。SIGL将检测恶意软件的安装过程作为主要研究目标，采用词向量编码Word2Vec对溯源图节点信息进行初步编码，随后使用GraphLSTM聚合节点的邻居属性传播节点属性的方式，使得SHADEWATCHER无法实现在线检测。

为了实现大规模日志数据的在线处理，香港大学的杨帆等人提出的PROGRAPHER[45]系统将流式溯源图分割成按时间排序的快照，并使用效率更高的Graph2vec对每个快照中的节点进行编码，最终分类出异常的节点信息。然而，这种快照排序的方式使得系统需要频繁进行磁盘与内存的调度，导致响应速度较慢。Rehman等人观察到，绝大多数系统实体在每次系统执行期间执行相同的一组活动，因此提出了FLASH[29]系统，在离线阶段通过存储这些节点的图神经网络（Graph Neural Network，GNN）向量，使得在运行时约80%的时间内可以避免直接使用GNN编码，从而有效降低系统的性能开销。另外，由于溯源图规模大，标记训练数据较为困难，Jiang等人提出的ORTHRUS[48]使用自构建的自监督神经网络模型进行训练，避免了数据标注。

现有的GNN技术在高效学习节点特征方面存在显著挑战。当前大多数基于节点或边特征的入侵检测系统采用GNN来捕捉这些特征，但由于GNN在处理大规模图数据时效率较低，这些系统难以实现快速响应的实时检测。此外，这些检测系统生成的细粒度告警信息通常缺乏对攻击链的因果分析，导致其可解释性不足，无法提供有效的攻击溯源和响应建议。

洞察3：基于节点/边的检测方法能够提供细粒度告警信息，缓解审计人员过滤无关信息的开销。

4.3　基于子图特征学习的入侵检测

在溯源图中，一个连通子图可以表示完整的入侵流程，基于子图进行攻击表示和检测是比较直观的想法。特别是攻击子图表示了潜在或进行中的攻

击行为。基于子图的溯源检测通常指在大规模溯源图中寻找并定位这些攻击图的过程。其中一大类工作尝试在训练阶段通过构建典型的系统子图行为模型来识别异常行为。如果检测阶段的子图行为明显偏离该模型，则会触发报警。

Manzoor 等人提出的 STREAMSPOT[24]设计了一种有效的流图检测方法，基于本地图遍历为每个节点创建标签，然后利用广度优先搜索算法生成能够表示子图结构的多个矩形，并使用相似性哈希将它们转换为固定长度的草图，最终将这些草图组合起来描述整个图。STREAMSPOT 的图编码主要考虑子图的结构信息，但忽略了子图中细粒度的节点、边属性以及时间属性，准确率较低。在此基础上，Han 等人提出的 UNICORN[51]使用了类似的方式创建标签，但其使用了类图核的方法生成能够表示子图结构和属性特征的直方图，并使用局部敏感哈希将其降维为固定长度的低维向量。与 STREAMSPOT 不同，UNICORN 考虑了子图的边属性和时间属性，并且支持大规模溯源计算。然而，UNICORN 的流图本质上是通过快照维护的，因此可能会出现检测延迟现象。上述工作检测的粒度都较为粗糙，结果难以解释和分析，无法满足实际使用中响应处置的需求。

为了提升结果的可解释性，Cheng 等人提出的 KARIOS[54]系统实现了一种基于时间窗口划分子图的异常检测方法。在一个时间窗口内，KARIOS 首先对每个节点特征进行编码，然后使用时态图网络（Temporal Graph Networks，TGN）将每个节点边缘周围的邻域结构和邻域中节点的状态作为新的向量，然后识别子图内一组可疑节点。KARIOS 使用这些可疑节点在不同子图下进行关联分析，增强了告警的解释性，以便实现快速的攻击调查。Altinisik 等人提出的 PROVG-SEARCHER[53]则采用了子图表示学习的方式对溯源图中的子图进行搜索，先利用定制的图分区方案将大规模的溯源图分割为 k-hop 的邻居图（ego-graph），并将分割后的子图向量化并进行子图相似图计算。浙江工业大学的吕明琪等人提出的 TREC[56]将采样的子图与战术、技术与过程（TTP）技术进行关联，实现 APT 攻击技术的映射。需要指出的是，无论采用什么分割算法，都很难准确地将攻击行为从正常行为中分割出来（否则该过程本身就完成了入侵检测任务）。因此，基于分割的方法必然会受到正常数据的干扰，或者将一个攻击分割到不同子图，从而损失检测的精度，而攻击者也可能利用这些问题实现针对性的对抗性攻击。

挑战 4：攻击图可以高效、准确地描述入侵流程。但如何从大规模溯源图中准确划分出攻击图是一个挑战。

4.4 基于路径特征学习的入侵检测

网络空间中的攻击行动通常具有高度的上下文关联性，通过一系列连续行为的执行以实现攻击目标（如图 4 中攻击者先进行木马工具传输，然后利用工具实现后续攻击目标）。这些连续的行为可以表示为一条攻击行为路径。根据该路径对攻击行为进行建模也是比较自然的想法。

相比于基于子图的检测而言，基于路径的检测算法更容易跟踪多步骤的长攻击链，且不容易被周围的正常行为干扰，在检测精度和效率上都有明显优势。华中科技大学的谢雨来等人提出的 PAGODA[57]和 P-GAUSSIAN[58]系统最早利用事件发生频率信息的方法进行攻击检测，有效提升了溯源分析的效率。Hassan 等人提出的 NODOZE[31]系统采用了类似的方法来筛选攻击相关因果路径。在其工作基础上，Wang 等人提出了 PROVDETECTOR[49]系统，该系统通过时间窗口改进了事件频次的统计方法，接着计算从溯源图中提取可疑路径并利用 Doc2Vec 进行编码，最后通过局部离群因子（Local Outlier Factor）来识别异常路径。然而，PROVDETECTOR 的计算需要离线加载整个溯源图，因此无法支持实时大规模的溯源图计算。

同时，基于溯源图的方法可以比较自然地与溯源分析经典的标签传播框架整合，实现日志流上的在线分析，大大降低数据缓存的开销和检测的延迟，提升检测效率。然而，现有的基于路径特征的入侵检测方法在解构溯源图上所有路径时面临着路径爆炸这一 NP 完全问题，需要在高效路径搜索和避免遗漏关键信息之间找到平衡；同时需要大量内存来缓存计算中间结果，增加了系统资源的消耗。

洞察 4：基于路径的检测方法能够关联到长时期系统实体间的依赖关系，具有更好的准确率和鲁棒性。同时，此类方法天然契合溯源分析中经典的基于标签传播的计算框架，能够实现更快的响应和更低的计算开销。

5 对抗性的攻击手段分析

对抗性是网络安全问题的基本属性之一。攻击

方设计攻击方案时必然会将详细了解防御方的能力纳入考虑，并针对性设计攻击方案。密码学中的“柯克霍夫原则①”也适用于入侵检测系统的设计。防御方应假设攻击者知道检测系统的算法设计，并对潜在的对抗性攻击进行分析和预防。同时，防御系统设计应该尽可能简洁，以避免复杂性导致的漏洞。随着基于溯源图的入侵检测技术研究展开，针对其设计缺陷的针对性攻击技术研究相关工作也得到了一定的关注[59]。本文将针对性的攻击手段分为两类进行讨论：针对检测流程的通用对抗性攻击手段以及针对学习技术引入新风险的对抗性攻击手段。

5.1 针对检测流程的通用对抗性攻击手段

高级持续性威胁具有高隐蔽性、长攻击周期等特点。在主机生成的海量审计日志中寻找有害事件无异于大海捞针。为了提高检测取证效率，一种常见的方法是通过“数据过滤→检测模块→告警过滤”三个阶段的串联使用来减少计算开销和误报率。然而，由于这些阶段紧密相连，攻击者往往可以通过针对某一特定阶段实施对抗攻击，从而干扰整个检测过程。

(1)数据过滤阶段。单台主机每天会生成大量日志，可达千兆字节，给数据存储和管理带来了严峻的挑战。由于防御方需要筛选无关的日志数据来寻找真正的证据，这会给检测带来“大海捞针”的问题。快速定位关键的日志信息是一个广泛研究的问题。最近的研究显示，常见的数据过滤手段，包括图压缩、语义修剪、信息流保存、因果关系近似等，可以一定程度上减少数据量[11,60-64]。这些手段期望过滤掉不会在检测中使用的正常事件，删除冗余的因果事件、重复的系统调用事件，甚至接受一些准确性损失以换取空间效率，从而缓解因大规模检测分析引发的成本问题。

然而，数据过滤算法无法准确保证去掉的数据都是正常的，攻击者可以利用该机制让过滤模块将攻击相关事件误判为正常事件并删去，从而规避后续的检测与分析。例如，攻击者可能利用系统中的软件更新通道下载恶意软件，并通过将其恶意活动伪装成与系统中常见的正常操作(如软件更新、系统定期维护任务、Git文件同步等)高度相似的行为，从而导致日志过滤算法误将这些恶意事件识别为正常事件并加以过滤。

(2)检测阶段。检测阶段的手段主要包括基于启发式的检测和基于异常的检测，基于启发式的检测系统根据已有的威胁知识提前定义知识库，如果系统中出现了相似的行为则会触发警报。Milajerdi等人提出的POIROT[14]系统构建了一组与高级持续性威胁相关的查询图，并利用定义的图匹配算法识别溯源图中的攻击行为。基于异常的检测方面，以复旦大学的贾子安等人构建MAGIC系统[30]为例，其将整个溯源图作为输入，然后使用图神经网络对每个系统活动实体进行编码以找出异常的离群实体。然而，一些隐蔽的攻击手段，包括离地攻击、无文件攻击等[1,65]，攻击者通过滥用良性应用程序来避免显性的异常活动，从而绕过检测系统。

(3)告警过滤阶段。现有的威胁检测系统容易出现高误报率，导致威胁警报疲劳问题。因此，在告警过滤阶段通常使用关联和聚类警报这两种手段缓解此问题，以减少需要调查的事件总数。一方面，一些研究工作查找与同一攻击相关的重复警报并将这些警报融合在一起，Pei等人提出的HERCULE系统[66]使用Louvain社区检测方法来发现异构审计日志中的攻击社区，然后利用这些攻击社区作为基础来得出威胁警报之间的相关性。然而，攻击者可以利用这些易于被减少的调查事件发动攻击，从而避免被分析人员调查取证。另一方面，为了进一步缓解威胁警报疲劳，分析人员通常会优先调查高严重性警报，而将低严重性警报放在后面。因此，攻击者可以将严重的威胁行为伪装成低严重性警报，使得分析人员无法在第一时间分析与处置它们。

洞察5： 入侵检测系统中的设计应尽可能保持简洁，额外的模块将向攻击者暴露更多的攻击面，影响整体系统面对对抗性攻击时的鲁棒性。尤其是数据过滤等可能会移除数据的模块的引入应保持谨慎。

5.2 针对学习技术引入新风险的对抗性攻击手段

近几年的研究成果已经证明了机器学习在溯源检测中应用的有效性。然而，机器学习方法本身存在的对抗攻击风险[67]，也被引入到基于学习的溯源检测中。目前，针对基于机器学习的溯源检测，学术界已经提出了基于干扰异质图的结构和属性来影响学习过程中的编码效果等对抗性攻击手段。

5.2.1 针对学习技术引入新风险的对抗性攻击手段

基于第4节中介绍的溯源图分析基础概念，本章节深入探讨了针对机器学习模型的对抗手段。具

① kerckhoff：https://www.petitcolas.net/kerckhoffs/index.html.

体而言，攻击者通过在溯源图G上施加特定的扰动操作集S来干扰模型的学习过程。这些扰动会影响图G的分解方式，将其细分为更小单元如序列、节点、路径或子图特征，并进一步影响这些特征的结构，从而达到绕过检测的目的。为形式化这一过程，本文提出以下公式化表达：

$$\eta:(G,S)\rightarrow z',\ G\in\{G_{\text{seq}},G_n,G_p,G_g\} \quad (2)$$

$$\varphi(z')<\alpha,\ |S|\leqslant\Delta \quad (3)$$

其中，α代表模型的告警阈值，当预测值大于告警阈值时，会触发系统报警。$|S|$表示扰动的幅度，而Δ表示扰动的最大允许范围，它衡量了检测模型的鲁棒性，Δ越小，说明检测模型的鲁棒性越高，这意味着攻击者能够实施的干扰行为种类和数量受限，使得绕过检测变得更加困难而不易被防御方发现。

许多基于机器学习的检测系统通过学习溯源图中正常行为的模式来检测异常。例如，一些检测系统[29-30,43]采用GNN对图结构中的实体进行编码，然后通过聚类等方法对这些编码后的节点进行预测。如图6所示的案例具体描述了一次横向移动攻击的过程：在受害者未察觉的情况下，攻击者首先向服务器上的nginx进程植入了后门。接着，被感染的nginx进程启动了一个远程连接工具，即test进程。一旦攻击者获得了对test进程的控制权，便利用它连接到网络地192.113.144.28，从而获取进一步访问内部服务器的权限。由于test进程的行为显著偏离了正常的活动范围，因此通常会被上述基于GNN的溯源检测系统标记为异常实体。

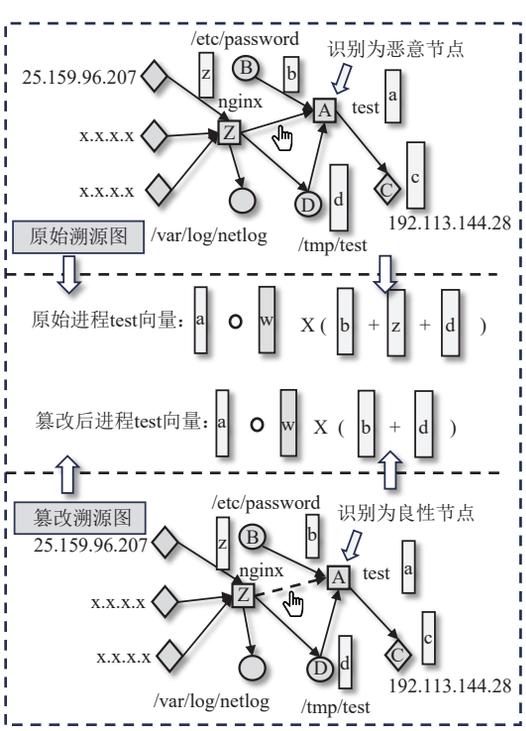

图6 逃逸攻击案例

然而，攻击者可以采取措施规避检测。例如，通过修改/tmp/test文件间接控制test进程的方式，攻击者能够改变该进程的行为特征以及其周围环境的活动模式。这不仅改变了涉及节点的编码值，还可能导致检测系统错误地将原本恶意的test进程分类为良性实体。这种对抗性操作凸显了现有基于机器学习的检测系统在面对精心设计的扰动时可能存在的脆弱性，并强调了开发更强大防御机制的重要性。

5.2.2 训练阶段的对抗性攻击手段

在训练基于机器学习的异常检测系统时，对抗性攻击手段通常针对这些系统的脆弱点，即通过篡改训练数据来破坏模型的有效性。异常检测方法依赖于识别那些与历史典型良性行为显著偏离的结构，以实现对高级持续性威胁的检测。然而，在实际应用中，攻击者可能采用隐蔽的数据投毒策略来污染历史审计数据，从而干扰检测系统的训练过程，使模型错误地将恶意行为误判为良性行为。

图7展示了在OpTC数据集中实施投毒攻击的一个案例。在原始的溯源图训练数据中，进程thunderbird仅连接内部网络，且与这些网络通信产生的文件都是相应的临时文件，例如事件<thunderbird，connect，128.55.12.10>和<thunderbird，write，#/128.55.12.10-I/inbox>。为了使检测模型中毒并绕过溯源检测，攻击者精心构造了包含外部网络61.130.69.232与进程thunderbird交互的事件，以及后续thunderbird进程写入恶意文件/var/log/mail的事件，并将其作为“毒”数据注入到训练集中。当

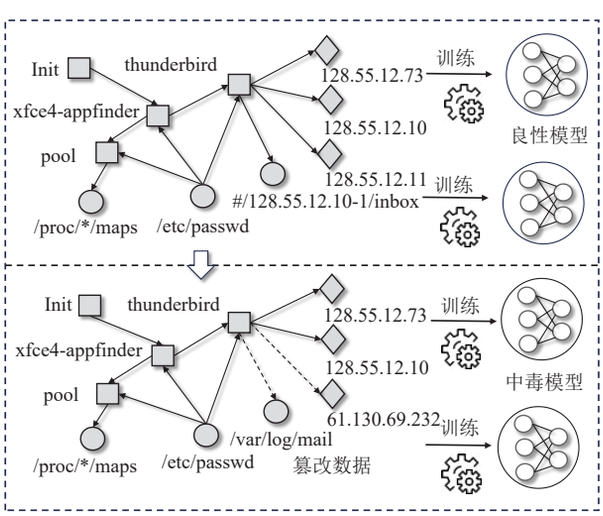

图7 针对PIDS的投毒攻击案例

使用这些被投毒的数据进行模型训练时，检测模型会错误地学习到这些恶意行为属于正常行为的一部分。因此，当攻击者真正发起攻击时，由于模型已经被投毒数据误导，它将无法有效地识别此类攻击行为，导致安全防护措施失效。

最近的几项工作对基于学习的溯源检测系统对抗投毒攻击的能力进行了评估。Han等人提出的SIGL系统利用恶意软件数据污染原始的训练数据集，以评估数据中毒对PIDS模型的影响。新加坡国立大学的曾俊等人提出的SHADEWATCHER系统和中国科学院大学的陈子军等人提出的KARIOS系统使用DARPA数据集中的一天攻击数据来评估数据中毒的影响。尽管他们的工作显示出一定的鲁棒性，但由于数据规模较小以及缺乏实际场景的攻击模拟，这些模型的评估具有误导性。数据中毒对基于学习的入侵检测系统依然构成潜在的严重威胁。

5.2.3　检测阶段的对抗性攻击手段

与影响模型训练的数据投毒攻击不同，逃逸攻击（Evasion Attacks）具有更广泛的潜在影响。尽管机器学习模型在多种应用领域的逃逸攻击方法已经得到了广泛研究，但在特征空间中的逃逸攻击与其在问题空间中的表现之间仍存在显著差距，尤其是在具有强问题空间依赖性的溯源检测领域。

具体来说，在检测阶段，攻击者在其控制的进程空间中精心策划隐蔽活动以修改待检测的溯源图结构信息，从而逃避检测。常见的干扰策略包括添加或删除进程与其他实体之间的关联，以及修改恶意图上实体的属性来生成对抗性攻击图。如图8所示，攻击者利用被感染的gtcache进程启动了一个新的profile进程，并通过该进程下载了恶意文件/var/log/mail。经过一段时间的潜伏后，mail进程在受害者系统中启动，实现了对目标主机的攻击。正常情况下，基于异常溯源检测的系统可以轻易识别出这种行为与历史正常行为之间的显著偏差。然而，在攻击者的精心策划下，通过修改mail文件的属性及周围活动的行为，使得这些恶意活动看起来与历史正常行为相似，从而使检测系统难以区分，最终实现攻击逃逸。

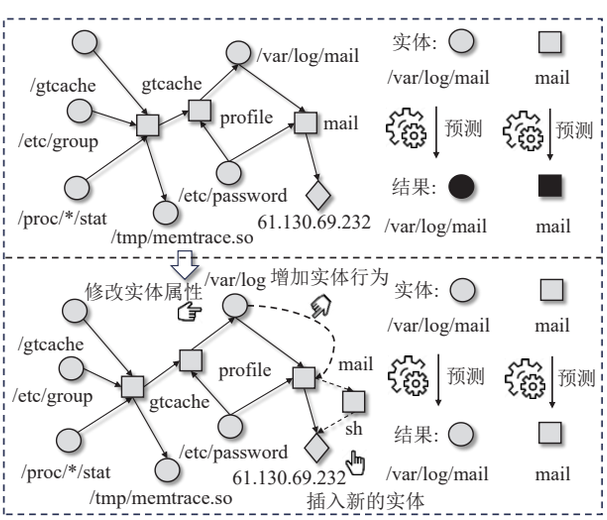

图8　针对PIDS的逃逸攻击

Mukherjee等人提出的PROVNINJA[68]框架展示了如何通过替换原始攻击中的明显异常事件为不明显事件来实现逃逸攻击。Goyal等人提出了一种方法，通过调整溯源图中恶意节点和良性节点邻域分布的相似性来迷惑检测系统[59]，Han等人利用伪装的恶意进程模仿良性进程测试了SIGL系统的鲁棒性，发现该系统在特定数据集上易受模仿攻击的影响。Rehman等人提出的FLASH系统虽然在检测Goyal等人模拟的攻击时表现出一定的有效性，但这些模拟攻击并非真实发生的逃逸攻击，因此对于攻击者能否使用公开信息和领域知识有效逃避基于机器学习的检测器的研究尚不充分。目前，大多数基于学习的入侵检测系统缺乏对防御对抗性攻击能力系统性的评估，因此难以从实证角度对这些其进行评估。表8结合理论分析和部分实验结果，总结了检测系统在应对两类对抗性攻击的能力，其中实心表示鲁棒性强，半实心表示有一定对抗能力，空心表示对抗能力弱。

表8　PIDS对抗逃逸攻击的鲁棒性

溯源入侵检测系统		攻击链替换[68]	插入良性事件[59]
基于节点	FLASH	●	未知
	SHADEWATCHER	未知	○
	SIGL	◐	◐
	R-CAID	●	未知
	ORTHRUS	●	未知
基于路径	PROVDECETOR	○	○
基于子图	STRAMSPOT	○	未知
	UNICORN	○	未知
	KAIROS	◐	◐

挑战5：机器学习算法在为溯源分析引入更强的能力的同时也引入了额外的风险。如何有针对性地优化机器学习算法，提升其在入侵检测时的鲁棒性是一个挑战。

5.3　对抗性攻击实施的难度

没有绝对安全的系统，评估对抗性攻击在受保

护主机上实施的难度是防御方考虑的重要指标之一。特别是在利用审计日志进行取证分析的溯源检测系统中，由于其具有极强的上下文关联性，因此防御方对防御知识的掌握程度以及攻击方能够控制的进程空间范围，会很大程度上影响攻击绕过难度，成为影响对抗性攻击成功与否的关键因素。攻击者对检测算法原理、模型参数、训练数据、标签和预测结果的了解程度是衡量防御边界的重要标准。基于机器学习的溯源检测系统中的对抗性攻击可以根据攻击者所掌握的信息量分为以下三类：(1)白盒攻击：攻击者全面了解检测系统的算法原理、模型参数、训练数据、标签和预测等全部信息。这种情况下，攻击者可以设计出高度针对性的攻击策略。(2)灰盒攻击：攻击者有限度地了解检测算法原理及其模型参数等信息。尽管不如白盒攻击那样全面，但仍然能根据已知信息调整攻击策略。(3)黑盒攻击：攻击者完全不了解检测算法和模型的具体信息，仅能访问模型输出的结果。在此情境下，攻击者需依赖于观察到的行为模式来推测并制定攻击策略。

此外，攻击者能够直接或间接控制的进程空间范围也是衡量其能否有效实施对抗攻击的一个重要因素。这包括攻击者在受害者主机上获得的权限级别、执行不同类型单步攻击的复杂度以及相应的时间开销等因素。最近的APT报告①②表明，许多APT组织倾向于使用0日漏洞或鱼叉式钓鱼邮件等隐蔽手段提升自身权限，并显著增加对系统实体空间的控制能力。对于敏感性较高的实体，直接操控可能立即引起防御方警觉，因此攻击者通常选择通过操控其他进程实体间接影响目标，以实现逃逸效果。寻找可用攻击路径，构造逃逸攻击对攻击者来说更加困难。

洞察6：任何防御手段都存在绕过风险，基于溯源的检测技术研究可以有效提升绕过难度，提升系统安全性。

6 机器学习用于溯源分析其他阶段

数据压缩与威胁情报提取等是溯源研究分析中的重要一环[21,69-70]。然而，由于高级持续性威胁具有高度复杂性和持久性，检测系统需要处理和分析海量日志数据，如何高效存储这些数据仍然是一个开放性的研究问题。同时，为了应对多样化的攻击手段，网络威胁情报(Cyber Threat Intelligence，CTI)的重要性日益凸显。然而，从大量非结构化的威胁情报报告(CTI Report)文本中人工提取关键知识不仅耗时，而且容易出错，因此自动化提取已成为研究的主流方向。目前，基于机器学习的数据压缩与威胁情报提取技术因其卓越的性能，正受到越来越多研究者的关注，本章将从多个维度分析现有方法的优势与局限性。

6.1 基于机器学习的存储优化

随着APT攻击持续时间延长和攻击手段复杂化，企业级环境下的系统日志和溯源图数据呈现爆炸性增长。如何高效存储这些日志以为溯源检测提供数据支撑，成为一个重要的挑战。目前，存储日志数据主要有两种类型的算法。一是检测无关的压缩算法。此类算法在设计之初以保留尽可能多的数据，不影响后续各种形式检测分析为压缩目标，如Xu等人[71]，提出的因果路径保留算法(Causality-Preserving Reduction，CPR)以保留完整的因果分析为目标，Hossain等人[11]提出了依赖关系保留的压缩算法(Dependence-Preserving Reduction，DPR)。上述算法需要保留尽可能多的语义，因此压缩倍率有限，分别为2.27×和4.6-19.1×左右。二是检测相关的压缩算法，结合后续的分析算法进行数据压缩。此类方法压缩倍率往往更高，如北京大学唐于涛等人提出的NODEMERGE[72]从程序执行模式角度切入，采用优化的FP-Growth算法分析程序访问规律以实现高效在线压缩。实验表明，相比原始数据，其存储空间压缩倍率可达28.3×。

而随着深度学习技术的发展，研究人员开始探索将神经网络应用于数据压缩。Ding等人提出的ELISE[61]，首次将深度神经网络(Deep Neural Network，DNN)引入日志压缩领域，通过字典机制和编码器训练实现高效无损压缩，相较于Gzip和DEEPZIP方法，ELISE平均可分别实现3倍和2倍的压缩率提升。同期，Fei等人提出的SEAL[73]提出了面向因果分析的查询友好压缩框架，创新性地结合了图结构压缩和时间局部性编码，在两个真实的数据集上分别实现了2.63-12.94×的压缩倍率。Ding等人提出的LEONARD[74]则将LSTM和校正表结合应用于溯源图存储，通过图结构解耦和LSTM序列建模，实现针对关联性查询任务无损压缩与无损解压，并且与关系数据库QuickStep和图

① APT CyberCriminal Campagin Collections. https://github.com/CyberMonitor/APT_CyberCriminal_Campagin_Collections

② Attack Wiki. https://attack.mitre.org/wiki/Groups

数据库Neo4j相比，LEONARD分别平均要减少17.43-25.9×的存储空间。这些压缩算法大多属于有损压缩，一定程度上会丢失语义，影响数据访问的效率，或限制数据访问的手段。因此需要配合定制的检测分析手段使用。

6.2 基于机器学习的查询优化

在溯源检测过程中，高效查询机制能够实现复杂路径的追踪与因果关系分析。例如，已知实体sshd已被注入，快速查询与其关联的事件。在早期的工作中[71-73,75]，溯源图主要通过关系型数据库或者图数据库进行查询。由于大型溯源图中包含大量冗余信息，导致查询的时间开销显著增加，因此关系数据库在查询溯源图时存在低效率问题。另外，关系数据库对图查询的支持较弱，难以表示数据点之间空间关系，且查询过程中往往涉及大量的I/O操作。具体来说，在进行反向溯源时，需要通过迭代查询和连接顶点表与边表来搜索关系，查询次数取决于任意顶点到目标顶点之间所有最短路径中的最长路径长度。对于规模更大的图，这种操作会导致更为沉重的I/O负担。相比之下，图数据库（如Neo4j）及其查询语言（如Cypher[76]和Gremlin[77]）在图的查询方面提供了更优的支持。图数据库以图结构形式存储节点和边，能够直接在图上进行查询操作，而无需执行复杂的表连接操作。

数据库的存储效率很大程度上取决于对冗余信息的识别能力，并且它们是否支持高效查询仍与存储格式密切相关。Gzip利用基于编码的压缩方法将整个图压缩为一个文件，并利用查询工具（如zcat、zgrep、zless和zdiff）实现有限的查询支持。然而，传统的编码工具多用于字符串操作，对图结构的查询支持相对较差。针对上述问题，Fei等人提出的SEAL[73]提出了面向因果分析的查询友好压缩框架，整体查询时间仅为未压缩数据的63.87%。Ding等人提出的LEONARD系统利用深度神经网络实现溯源图查询，LEONARD的核心思想是将溯源图转换为数值向量，并利用深度神经网络进行查询。这种方法利用DNN模型替代传统的数据库索引，与关系数据库QuickStep和图数据库Neo4j相比，LEONARD提高了99.6%的查询速度，实现了高效的图查询和溯源分析。

展望2：溯源图中存在大量冗余、无关信息，通过合理利用机器学习算法可以实现对关键信息的高效的筛选、总结与压缩，有效提升溯源数据存储、查询和处理效率。

6.3 基于机器学习的威胁情报提取

威胁情报是有关现有或新兴网络威胁的基于证据的知识，可以促进应对网络威胁的决策过程。例如，通过提供攻击者TTP的语义化描述，威胁情报可以为溯源检测构建细粒度的威胁知识图谱，并利用子图匹配在溯源图上发现可疑的攻击子图。为了促进基于威胁情报共享的知识交换和管理，安全社区采用了标准化开放格式（例如OpenIoC、STIX和CybOX）来描述攻击指示器（IoC）和其他高层的攻击知识。威胁情报提取通过提取关键IoC和攻击行为，帮助识别威胁源头、构建攻击链条和绘制行为图，提升溯源检测的精准性和效率。然而，由于大多数威胁情报报告以自然语言的非正式格式编写，提取结构化攻击行为需要分析非结构化文本中的语义，并且自然语言文本中的大量细微描述差别限制了提取的准确性和提取的效率。

如图9展示了一个名为“Frankenstein”的真实APT攻击活动①，该活动得名于攻击者能够将多种独立技术拼凑组合的能力。如图所示，该活动包含四种攻击技术：网络钓鱼邮件、用户执行、漏洞利用以及启动项自启动。每项技术都涉及多个实体和依赖关系，用以实现一个或多个战术攻击目标，呈现了一个由多个原子技术组成的典型多阶段攻击活动。图9（A）表示手工生成的真实攻击图，图9（B-G）则展示了不同工作自动化提取技术的效果图。Milajerdi等人提出的POIROT[14]利用手动提取的广义攻击查询图对系统日志构建的溯源图进行检测，验证了威胁情报检测的有效性。然而，从无标注的文本中手动提取攻击相关信息既费力又容易出错，阻碍了威胁情报的实际应用。

针对该问题，自动化技术逐渐成为主流研究方向。Husari等人提出的TTPDRILL[78]使用自然语言处理（NLP）中的词性标注技术自动解析CTI报告，从中提取候选威胁行为，接着TTPDRILL结合依赖解析和启发式方法，将从文档中提取的威胁操作映射至杀伤链阶段。Zhu等人提出的CHAINSMITH[79]采用四阶段模型定义攻击活动，通过训练多类分类器提取IoC，并进一步将其分类到不同阶段。浙江大学的李振源等人提出ATTACKG[2]，结合机器学习与规则匹配实现实体提取和关系识别，并通过改进的图对齐算法，将攻击

① Frankenstein campaign：https://blog.talosintelligence.com/2019/06/frankenstein-campaign.html.

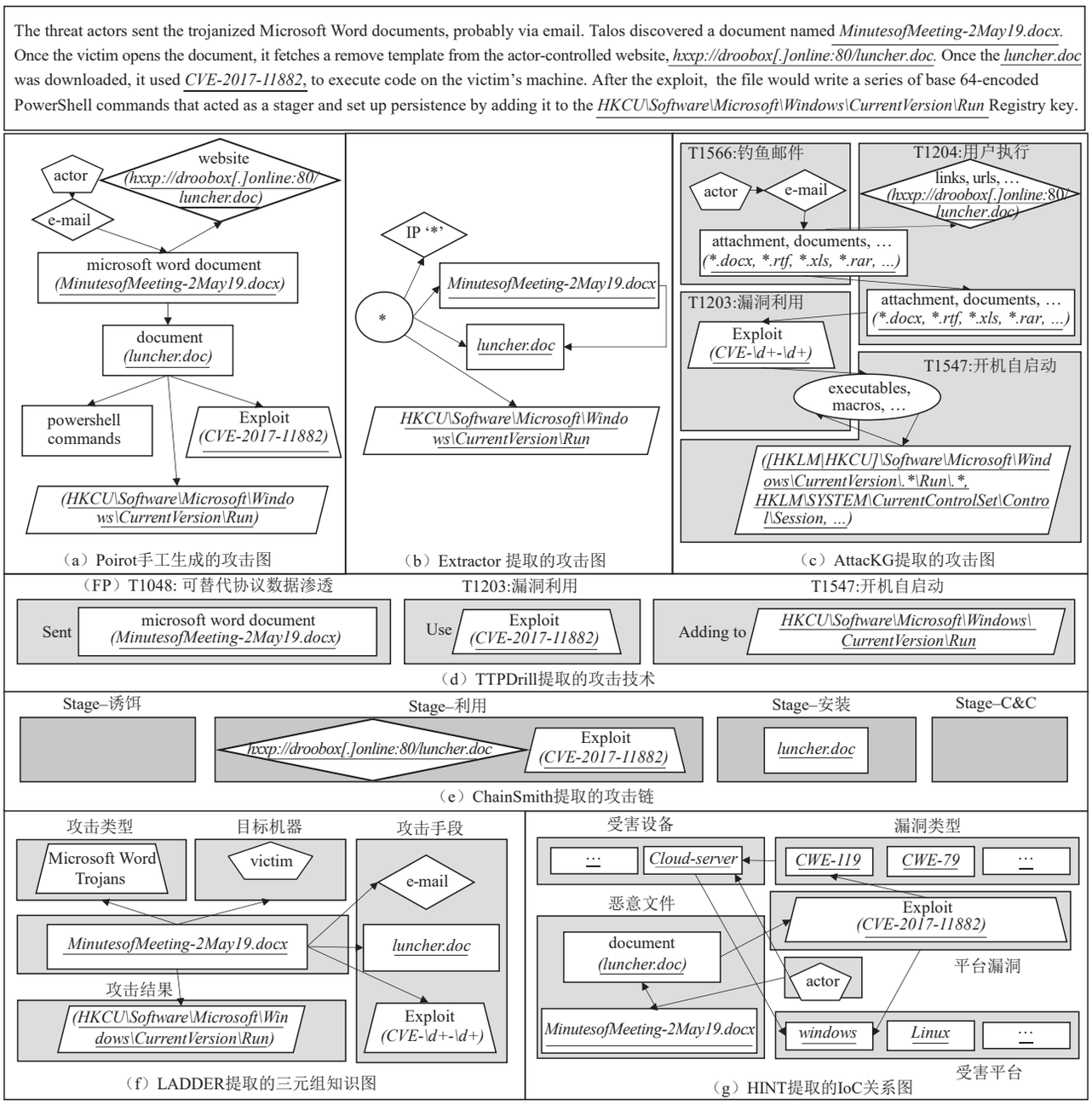

图9 威胁情报知识提取效果对比

图中的技术模板与MITRE ATT&CK ID进行映射，从而构建技术知识图。

与此同时，一些研究尝试摆脱对预定义静态模式的依赖，通过定制的自然语言处理技术直接从文本中建模攻击行为为攻击图。北京航空航天大学的赵等人提出了HINTI方法[80]，通过基于多粒度注意力的IoC识别技术提高IoC提取的准确性，并利用异构信息网络（HIN）建模IoC之间的相互依赖关系。Peng等人提出的THREATRAPTOR[81]提出了一种无监督、轻量级且高效的NLP管道，从非结构化文本中提取结构化威胁行为，同时结合IoC保护和依赖解析技术提取IoC关系。Satvat等人提出的EXTRACTOR[82]通过词性标注、依存句法分析、语义依存分析等NLP技术对规范化、解析、总结与知识图谱构建进行优化，采用基于BERT和BiLSTM的深度学习方法减少冗余信息，提升了从威胁报告中提取攻击图的效率。然而，其管道中并未包含攻击模式提取组件。相比之下，LADDER[83]提出应结合使用更强大的威胁情报信号（如攻击模式）与IoC，并利用TTPClassifier方法从CTI报告中提取相关的TTP，将其重组为知识图（Knowledge Graph），进一步提升对恶意行为分析的支持能力。

洞察7：自然语言处理技术能高效提取非结构化的攻击报告中的知识，并结构化表示，有效提升威胁分析效率。

7 讨论、总结与展望

7.1 理想的入侵检测系统设计

理想的溯源检测系统应当兼顾高准确率、鲁棒性、效率以及工程实践的可行性与可用性。其核心设计应聚焦以下几个关键维度：

(1)资源效率与高精度检测能力。在严格的资源约束条件下(例如内存占用<20 MB/单主机[37])，系统需实现对跨天或跨周复杂攻击的持续监测。系统应在不依赖全量历史日志存储的前提下，完整保留攻击路径的关键语义信息，并支持每秒处理万级日志事件的实时流式过滤和动态优先级排序功能，以减少冗余数据对资源的占用。此外，系统还需特别关注隐蔽攻击模式(如慢速扫描、无文件攻击、离地攻击等)的检测，避免因特征隐匿导致的检测盲区。

(2)复杂攻击追溯与可解释性能力。系统需要能够构建覆盖跨主机和服务的攻击因果链，通过多维度实体关联重构从初始入侵点到最终目标的攻击图谱。告警信息应采用双层表达结构：技术层提供精确的攻击手段描述(如进程调用链、文件操作序列等)，战术层则语义化地呈现攻击意图(如数据窃取、权限提升等)。借助因果推理引擎实现告警精准聚合，确保每次警报仅关联同一攻击链中的必要事件，同时自动抑制重复或低优先级告警，使日均有效告警量达到行业使用预期。

(3)开放世界场景下的适应性能力。系统必须能够识别未见过的新型攻击模式(如0日攻击)，无须依赖预定义规则或特征库。它还应该支持跨异构环境(云、边缘、容器)的行为模式泛化，防止因环境和背景行为差异导致的检测失效。此外，在环境行为场景发生变化时，面对概念漂移问题，系统需具备在线更新能力，以快速适应新威胁并避免因为新行为和基线模型不一致导致的误报问题。

7.2 主要的差距和展望

尽管基于学习的溯源图检测方案已取得显著进展，但在实际应用中仍面临如下瓶颈：

(1)内存占用过高。现有研究往往忽视了对内存消耗的有效管理，尤其是在训练和检测阶段。文献分析显示，当前系统的内存优化措施不足或方案不完整，主要原因是系统运行期间将大量溯源数据加载至内存中。例如，PROVDETECTOR和KAIROS系统分别在检测和训练阶段遭遇内存瓶颈，限制了它们在大规模审计数据上的应用。

(2)可解释性不足。相较于基于规则的系统，基于学习的检测系统由于机器学习模型的黑盒特性，通常缺乏足够的可解释性。虽然像FLASH和MAGIC这样的系统实现了细粒度告警，但未能充分提供攻击行为的上下文关联分析或将原始溯源图抽象至易于理解的技术策略层级。

(3)适应能力缺失。现实世界的生产环境比实验室场景更为复杂，要求检测系统具备应对持续概念漂移的能力。然而，现有系统在面对新型攻击或未知良性行为时容易产生误判，即使一些改进方案试图扩展基线学习范围，但由于采用了简单的泛化匹配策略，在处理未表征实体时会导致较高的误报率。

现有方案通过编码器实现实体信息的高维至低维嵌入表示，并借助基线模型对比分析威胁实体，使得攻击者难以利用慢速攻击、持久性攻击等隐蔽攻击在主机上进行长期渗透。然而，在面对集成了无服务器和Web应用程序等多源数据时，研究者们难以设计出一个通用的方法在降低内存占用和确保低延迟响应的同时保证检测结果的准确性和可解释性，并且检测系统需要进一步平衡多源数据流吞吐量与处理速度。因此，在大规模流式溯源图上实现基于学习的高效溯源检测系统是未来的重要研究方向。

展望3：入侵检测是一个系统性问题。当前基于智能溯源分析的入侵检测各模块在部分属性上实现了比较理想的效果，但这些模块之间往往不能简单地组合到一起。如何合理组合使用这些模块，达到整体效果最优是一个开放性问题。

7.3 总 结

随着隐蔽威胁(如离地攻击、无文件攻击)的不断发展，基于大规模多源异构日志的溯源检测技术因其能够关联分析复杂攻击路径与意图而变得尤为重要，成为对抗高级持续性威胁的核心工具。本文探讨了机器学习在这一领域的应用挑战，从节点、路径、子图3个层面分析了各种检测方法的研究进展，并讨论了对抗性攻击的影响及未来方向，提出了一些潜在的研究方向，包括融合流式计算架构与图数据库技术、构建支持实时机器学习的高效溯源检测框架等。

参 考 文 献

[1] LI Z, CHEN Y, CHEN Q, et al. Effective and light-weight deobfuscation and semantic-aware attack detection for powerShell scripts//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS′19). London, UK, 1831-1847

[2] LI Z, ZENG J, CHEN Y, et al. Attackg: constructing technique knowledge graph from cyber threat intelligence reports//Proceedings of Computer Security-ESORICS 2022: 27th European Symposium on Research in Computer Security, Copenhagen, Denmark, 2022, 589-609

[3] XIONG C, LI Z, CHEN Y, et al. Generic, efficient, and effective deobfuscation and semantic-aware attack detection for powerShell scripts. Frontiers of Information Technology & Electronic Engineering. 2022. 23, 361-381

[4] SHEN X, LI Z, BURLEIGH G, et al. Decoding the mitre engenuity att&ck enterprise evaluation: an analysis of edr performance in real-world environments//Proceedings of the 19th ACM Asia Conference on Computer and Communications Security (ASIA CCS′24). Singapore, 2024: 96-111

[5] LI Z, WEI Y, SHEN X, et al. Marlin: Knowledge driven analysis of provenance graphs for efficient and robust detection of cyber attacks arXiv: 2403.12541, 2024

[6] Zhang Fan, Zhao Xin-Jie, Guo Shi-Ze. Enigma countermeasure: the development direction of highly concealed threat Insight in cyberspace. Communications of the CCF, 2023, 19(3): 97-104. (in Chinese)
(张帆,赵新杰,郭世泽. 密态对抗——网络空间高隐蔽威胁透视的发展方向. 中国计算机学会通讯,2023,19(3):97-104)

[7] Zhang J, Zhang C, Xuan JF, et al. Recent progress in program analysis. Journal of Software, 2019, 30(1): 80-109 (in Chinese)
(张健, 张超, 玄跻峰, 等. 程序分析研究进展. 软件学报,2019, 30(1): 80-109)

[8] Fan Zu-Wei, Zhang Shun-Liang, Zhao Hong-Ce. Survey on deep learning based malicious encrypted traffic detection and adversarial techniques. Journal of Cyber Security, to appear (in Chinese)
(樊祖薇, 张顺亮, 赵泓策. 基于深度学习的恶意加密流量检测及对抗技术综述. 信息安全学报, 已采用)

[9] Liao Xiang-Ke, Li Shan-Shan, Dong Weiet al. Survey on log research of large scale software system. Journal of Software, 2016, 27(8):1934-1947. (in Chinese)
(廖湘科, 李姗姗, 董威, 等. 大规模软件系统日志研究综述. 软件学报, 2016, 27(8):1934-1947)

[10] Ji Tian-Tian, Fang Bin-Xing, Cui Xianget al. Research on deep learning-powered malware attack and defense techniques. Chinese Journal of Computers, 2021, 44(4): 669-695. (in Chinese)
(冀甜甜, 方滨兴, 崔翔, 等. 深度学习赋能的恶意代码攻防研究进展. 计算机学报, 2021, 44(4): 669-695)

[11] HOSSAIN M N, WANG J, SEKAR R, et al. Dependence-preserving data compaction for scalable forensic analysis//Proceedings of the 27th USENIX Conference on Security Symposium (SEC′18). Baltimore, USA, 2018: 1723-1740

[12] GAO P, XIAO X, LI D, et al. Saql: a streambased query system for real-time abnormal system behavior detection//Proceedings of the 27th USENIX Conference on Security Symposium (SEC′18). Baltimore, USA, 2018: 639-656

[13] GAO P, XIAO X, LI Z, et al. Aiql: enabling efficient attack investigation from system monitoring data//Proceedings of the 2018 USENIX Conference on Usenix Annual Technical Conference (USENIX ATC′18). Boston, USA, 2018: 113-126

[14] MILAJERDI S M, ESHETE B, GJOMEMO R, et al. Poirot: aligning attack behavior with kernel audit records for cyber threat hunting//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS′19). London, UK, 2019: 1795-1812.

[15] LI Z, WEI Y, SHEN X, et al. Tags: realtime intrusion detection with tag-propagation-based provenance graph alignment on streaming events. arXiv:2403.12541, 2024

[16] MILAJERDI S M, GJOMEMO R, ESHETE B, et al. Holmes: real-time apt detection through correlation of suspicious information flows//Proceedings of 2019 IEEE Symposium on Security and Privacy (SP), San Francisco, USA, 2019: 1137-1152

[17] HOSSAIN M N, MILAJERDI S M, WANG J, et al. Sleuth: real-time attack scenario reconstruction from cots audit data//Proceedings of the 26th USENIX Conference on Security Symposium (SEC′17). Vancouver, Canada, USA, 2017: 487-504

[18] SCHEIN A I, POPESCUL A, UNGAR L H, et al. Methods and metrics for cold-start recommendations//Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR ′02). Tampere, Finland, 2002: 253-260

[19] JIANG J, HAN C, ZHAO W X, et al. Pdformer: propagation delay-aware dynamic long-range transformer for traffic flow prediction//Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence (AAAI′23/IAAI′23/EAAI′23), 2023, 3: Article 487, 4365-4373

[20] KRASKA T, BEUTEL A, CHI E H, et al. The case for learned index structures. //Proceedings of the 2018 International Conference on Management of Data (SIGMOD′18). Houston, USA, 2018: 489-504

[21] INAM M, CHEN Y, GOYAL A, et al. Sok: History is a vast early warning system: auditing the provenance of system intrusions//Proceedings of the 2023 IEEE Symposium on Security and Privacy (SP), San Francisco, USA, 2023:2620-2638

[22] WANG L, SHEN X, LI W, et al. Incorporating gradients to rules: towards lightweight, adaptive provenance-based intrusion detection//Proceedings of Network and Distributed System Security (NDSS) Symposium 2025, San Diego, USA, 2025:1-18

[23] ZENG J, WANG X, LIU J, et al. Shadewatcher: recommendation-guided cyber threat analysis using system audit records//Proceedings of 2022 IEEE Symposium on Security and

Privacy (SP), San Francisco, USA, 2022: 489-506

[24] MANZOOR E, MILAJERDI S M, AKOGLU L. Fast memory-efficient anomaly detection in streaming heterogeneous graphs//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD′16). San Francisco, USA, 2016: 1035-1044

[25] LI Z, CHEN Q A, YANG R, et al. Threat detection and investigation with system-level provenance graphs: a survey. Computers & Security, 2021, 106: 102282

[26] ZIPPERLE M, GOTTWALT F, CHANG E, et al. Provenance-based intrusion detection systems: a survey. ACM Computing Surveys, 2022, 55(7), article 135

[27] MOREAU L, FREIRE J, FUTRELLE J, et al. The open provenance model: an overview//Proceedings of Provenance and annotation of Data and Processes: Second International Provenance and Annotation Workshop, IPAW 2008, Salt Lake City, USA,2008: 323-326

[28] MUNISWAMY-REDDY K K, SELTZER M. Provenance as firstclass cloud data. ACM SIGOPS Operating Systems Review, 2010, 43(4): 11-16

[29] REHMAN M U, AHMADI H, HASSAN W U. Flash: a comprehensive approach to intrusion detection via provenance graph representation learning//Proceedings of 2024 IEEE Symposium on Security and Privacy (SP) San Francisco. USA, 2024: 139-139

[30] ZIAN JIA Y N, Yun Xiong. Magic: detecting advanced persistent threats via masked graph representation learning//Proceedings of the 33rd USENIX Conference on Security Symposium (SEC′24. FranciscoSan, USA, 2024: 1-18

[31] HASSAN W U, GUO S, LI D, et al. Nodoze: combatting threat alert fatigue with automated provenance triage//Proceedings of Network and Distributed System Security (NDSS) Symposium, San Diego, USA, 2019: 1-15

[32] FANG P, GAO P, LIU C, et al. Back-Propagating system dependency impact for attack investigation//Proceedings of 31st USENIX Security Symposium (USENIX Security 22). Boston, USA, 2022: 2461-2478

[33] LI S, DONG F, XIAO X, et al. Nodlink: An online system for fine-grained apt attack detection and investigation.//Proceedings of Network and Distributed System Security (NDSS) Symposium 2024, San Diego, USA, 2024:1-18

[34] LIU J, INAM M A, GOYAL A, et al. What we talk about when we talk about logs: understanding the effects of dataset quality on endpoint threat detection research//Proceedings of 2025 IEEE Symposium on Security and Privacy (SP). San Francisco, USA, 2025: 112-129

[35] YAO ZHU Y W S J, LiZhenyuan. The case for learned provenance-based system behavior baseline//Proceedings of International Conference on Machine Learning (ICML). Vancouver, Canada, 2025:1-15

[36] LANDAUER M, SKOPIK F, FRANK M, et al. Maintainable log datasets for evaluation of intrusion detection systems. IEEE Transactions on Dependable and Secure Computing, 2023, 20(4): 3466-3482

[37] DONG F, LI S, JIANG P, et al. Are we there yet? an industrial viewpoint on provenance-based endpoint detection and response tools [C]//Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS′23). Copenhagen, Denmark, 2023: 2396-2410

[38] DONG F, WANG L, NIE X, et al. Distdet: a cost-effective distributed cyber threat detection system//Proceedings of the 32nd USENIX Conference on Security Symposium. (SEC′23). Anaheim, USA, 2023: 6575-6592

[39] FANG Y, WANG K, LIN X, et al. Cohesive subgraph search over big heterogeneous information networks: applications, challenges, and solutions//Proceedings of the 2021 International Conference on Management of Data (SIGMOD′21). Virtual Event China, 2021: 2829-2838

[40] DU M, LI F, ZHENG G, et al. Deeplog: anomaly detection and diagnosis from system logs through deep learning//Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS′17). Dallas, USA, 2017: 1285-1298

[41] LIU F, WEN Y, ZHANG D, et al. Log2vec: a heterogeneous graph embedding based approach for detecting cyber threats within enterprise//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security (CCS′19). London, UK, 2019: 1777-1794

[42] ALSAHEEL A, NAN Y, MA S, et al. Atlas: a sequence-based learning approach for attack investigation.//Proceedings of the 30th USENIX Security Symposium (USENIX Security 21). Virtual, 2021: 3005-3022

[43] HAN X, YU X, PASQUIER T, et al. Sigl: securing software installations through deep graph learning//Proceedings of the 30th USENIX Security Symposium (USENIX Security 21). Virtual, 2021: 2345-2362

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

[45] YANG F, XU J, XIONG C, et al. Prograoher: an anomaly detection system based on provenance graph embedding//Proceedings of the 32nd USENIX Conference on Security Symposium (SEC′23). Anaheim, USA, 2023: 4355-4372

[46] REHA J, LOVISOTTO G, RUSSO M, et al. Anomaly detection in continuous-time temporal provenance graphs//Proceedings of Temporal Graph Learning Workshop@NeurIPS 2023. Louisiana, USA, 2023:1-16

[47] GOYAL A, WANG G, BATES A. R-caid: embedding root cause analysis within provenance-based intrusion detection//Proceedings of 2024 IEEE Symposium on Security and Privacy, San Francisco, USA, 2024: 257-257. Symposium

[48] JIANG B, BILOT234 T, MADHOUN NEL, et al. Orthrus: achieving high quality of attribution in provenance-based intrusion detection systems//Proceedings of the 34th USENIX Security Symposium (USENIX Security 25). Seattle, USA, 2025:1-20

[49] WANG Q, HASSAN W, LI D, et al. You are what you do: hunting stealthy malware via data provenance analysis//

Proceedings of Network and Distributed Systems Security (NDSS) Symposium 2020. San Diego, USA, 2020:1-17

[50] MENG L, XI R, LI Z, et al. Pg-aid: an anomaly-based intrusion detection method using provenance graph//Proceedings of 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Tianjin, China, 2024: 2522-2527

[51] HAN X, PASQUIER T F J, BATES A, et al. Unicorn: runtime provenance-based detector for advanced persistent threats//Proceedings of Network and Distributed Systems Security (NDSS) Symposium 2020. San Diego, USA, 2020:1-18

[52] CHEN T, DONG C, LV M, et al. Apt-kgl: an intelligent apt detection system based on threat knowledge and heterogeneous provenance graph learning. IEEE Transactions on Dependable and Secure Computing, 12(1), 2022, 1-15

[53] ALTINISIK E, DENIZ F, SENCAR H T. Provg searcher: a graph representation learning approach for efficient provenance graph search//Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security (CCS'23). Copenhagen, Denmark, 2023: 2247-2261

[54] CHENG Z, LV Q, LIANG J, et al. Kairos: practical intrusion detection and investigation using whole-system provenance.// Proceedings of the 2024 IEEE Symposium on Security and Privacy, San Francisco, USA, 2024: 3533-3551

[55] ALY A, IQBAL S, YOUSSEF A, et al. Megr-apt: a memory-efficient apt hunting system based on attack representation learning, IEEE Transactions on Information Forensics and Security, 2024, 19(6): 5257-5271

[56] LV M, GAO H, QIU X, et al. Trec: apt tactic/technique recognition via few-shot provenance subgraph learning// Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security (CCS'24). Salt Lake City, USA, 2024: 139-152

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

[58] XIE Y, WU Y, FENG D, et al. P-Gaussian: provenance-based gaussian distribution for detecting intrusion behavior variants using high efficient and real time memory databases. IEEE Transactions on Dependable and Secure Computing, 2019, 18 (6): 2658-2674

[59] GOYAL A, HAN X, WANG G, et al. Sometimes, you aren't what you do: mimicry attacks against provenance graph host intrusion detection systems//Proceedings of Network and Distributed System Security (NDSS) Symposium 2023. San Diego, USA, 2023:1-18

[60] XIE Y, MUNISWAMY-REDDY K K, FENG D, et al. Evaluation of a hybrid approach for efficient provenance storage. ACM Transactions on Storage, 2013, 9(4): 1-29

[61] DING H, YAN S, ZHAI J, et al. Elise: a storage efficient logging system powered by redundancy reduction and representation learning//Proceedings of the 30th USENIX Security Symposium. Virtual, 2021: 3023-3040

[62] BATES A, TIAN D, HERNANDEZ G, et al. Taming the costs of trustworthy provenance through policy reduction. ACM Transactions on Internet Technology, 2017, 17(4): 1-21

[63] HASSAN W U, LEMAY M, AGUSE N, et al. Towards scalable cluster auditing through grammatical inference over provenance graphs//Proceedings of the Network and Distributed Systems Security (NDSS) Symposium 2018. San Diego, USA, 2018:1-15

[64] MICHAEL N, MINK J, LIU J, et al. On the forensic validity of approximated audit logs//Proceedings of the 36th Annual Computer Security Applications Conference (ACSAC'20). Austin, USA, 2020: 189-202

[65] BARR-SMITH F, UGARTE-PEDRERO X, GRAZIANO M, et al. Survivalism: systematic analysis of windows malware living-off-the-land//Proceedings of the 2021 IEEE Symposium on Security and Privacy, San Francisco, USA, 2021: 1557-1574

[66] PEI K, GU Z, SALTAFORMAGGIO B, et al. Hercule: attack story reconstruction via community discovery on correlated log graph//Proceedings of the 32nd Annual Conference on Computer Security Applications (ACSAC'16). Los Angeles, USA, 2016: 583-595

[67] WEI J, YAXIN L, HAN X, et al. Adversarial attacks and defenses on graphs: a review, a tool, and empirical studies. arXiv: 2003.00653, 2020

[68] MUKHERJEE K, WIEDEMEIER J, WANG T, et al. Evading provenance-based ml detectors with adversarial system actions//Proceedings of the 32nd USENIX Conference on Security Symposium (SEC'23). Anaheim, USA, 2023: 1199-1216

[69] WANG W, ZHENG V W, YU H, et al. A survey of zero-shot learning: settings, methods, and applications. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-37

[70] SHIN B, LOWRY P B. A review and theoretical explanation of the 'cyberthreat-intelligence (cti) capability' that needs to be fostered in information security practitioners and how this can be accomplished. Computers & Security, 2020, 92: 101761

[71] XU Z, WU Z, LI Z, et al. High-fidelity data reduction for big data security dependency analyses//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS'16). Vienna, Austria, 2016: 504-516

[72] TANG Y, LI D, LI Z, et al. Nodemerge: template based efficient data reduction for big-data causality analysis//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS'18). Toronto, Canada, 2018: 1324-1337

[73] FEI P, LI Z. Seal: Storage-efficient causality analysis on enterprise logs with query-friendly compression //Proceedings of 30th USENIX Security Symposium (USENIX Security 21). Vancouver, Canada, 2021: 2987-3004

[74] DING H, ZHAI J, DENG D, et al. The case for learned provenance graph storage systems//Proceedings of 32nd USENIX Security Symposium (Usenix Security 23). Anaheim, USA, 2023: 3277-3294

[75] LEE K H, ZHANG X, XU D. LogGC: garbage Collecting Audit Log//Proceedings of the 2013 ACM SIGSAC

Conference on Computer & Communications Security (CCS'13). Berlin, Germany, 2013: 1005-1016

[76] FRANCIS N, GREEN A, GUAGLIARDO P, et al. Cypher: an evolving query language for property graphs//Proceedings of the 2018 International Conference on Management of Data. Houston, USA, 2018: 1433-1445

[77] RODRIGUEZ M A. The gremlin graph traversal machine and language (invited talk)// Proceedings of the 15th Symposium on Database Programming Languages. Pittsburgh, USA, 2015: 1-10

[78] HUSARI G, AL-SHAER E, AHMED M, et al. Ttpdrill: automatic and accurate extraction of threat actions from unstructured text of cti sources//Proceedings of the 33rd Annual Computer Security Applications Conference (ACSAC'17). Orlando, USA, 2017: 103-115

[79] ZHU Z, DUMITRAS T. Chainsmith: automatically learning the semantics of malicious campaigns by mining threat intelligence reports//Proceedings of 2018 IEEE European Symposium on Security and Privacy (EuroS&P). London, UK, 2018: 458-472

[80] ZHAO J, YAN Q, LIU X, et al. Cyber threat intelligence modeling based on heterogeneous graph convolutional network// Proceedings of the 23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020). San Sebastian, Spain, 2020: 241-256

[81] GAO P, SHAO F, LIU X, et al. Enabling efficient cyber threat hunting with cyber threat intelligence//Proceedings of the 37th IEEE International Conference on Data Engineering (ICDE). Chania, Greece, 2021: 193-204

[82] SATVAT K, GJOMEMO R, VENKATAKRISHNAN V. Extractor: extracting attack behavior from threat reports// Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P). Vienna, Austria, 2021: 598-615

[83] ALAM M T, BHUSAL D, PARK Y, et al. Looking beyond iocs: automatically extracting attack patterns from external cti// Proceedings of the 26th International Symposium on Research in Attacks, Intrusions and Defenses. Hong Kong, China, 2023: 92-108

LI Zhen-Yuan, Ph. D., assistant professor. His research interests include system security, intrusion detection and cyber threat analysis.

WEI Yang-Yang, M. S. candidate. His main research interests include provenance intrusion detection and defense against advanced persistent threats (APT).

WANG Zheng-Kai, M. S. candidate. His main research interests include provenance intrusion detection and defense against advanced persistent threats (APT).

JI Shou-Ling, Ph. D., professor. His current research interests include data-driven security and Al security.

Background

Detecting Advanced Persistent Threats are a critical research direction in cybersecurity. Since the U. S. Defense Advanced Research Projects Agency (DARPA) launched the "Transparent Computing" initiative in 2016, the research paradigm in this field has undergone significant transformation. This project introduces a foundational framework for achieving behavioral observability through systematic log collection and modeling, driving provenance intrusion detection research toward analyzing threat behaviors based on information flows and interaction relationships among system entities. This approach established theoretical foundations for constructing transparent system monitoring architectures.

With advancements in graph computing technologies, machine learning-based graph representation methods have demonstrated remarkable advantages in provenance detection. This paper systematically summarizes the research framework of provenance detection through three dimensions: algorithm design, defense mechanisms against adversarial attacks, and optimization of other related processes. By dissecting the fundamental theoretical framework of provenance detection, it emphasizes the representational advantages conferred by machine learning techniques to detection systems and the emerging security challenges they introduce. Ultimately, this paper reveals the existing technological gaps and implementation chasm between current academic research outcomes and the practical requirements of industrial-grade detection systems.

This work was supported by the "Pioneer" and "Leading Goose" R&D Program of Zhejiang (2024C03288, 2025C02263), the National Natural Science Foundation of China (NSFC) (No. 62402419), the National Key Research & Development Project of China (2023YFB3106800), the Ningbo Yongjiang Talent Programme and by the CCF-NSFocus "KunPeng" Research Fund.

Together, these programs serve the long-term goal of building China's cyberspace offensive, defensive, and countermeasure capabilities. This paper summarizes this latest progress in the field of intrusion detection based on intelligent provenance analysis to support the subsequent research.