基于溯源图的网络攻击调查研究综述

仇 晶[1,2],陈荣融[1],朱浩瑾[3],肖岩军[4],殷丽华[1],田志宏[1*]
(1. 广州大学网络空间安全学院,广东广州 510555;2. 鹏城实验室,广东深圳 518000;
3. 上海交通大学计算机科学与工程系,上海 200240;4. 绿盟科技集团股份有限公司,北京 100089)

摘 要: 网络攻击调查是实现主动防御、溯源反制的重要手段. 面向高隐蔽、强对抗的现代网络攻击,研究高效率、自动化攻击调查方法,提升己方快速响应复杂网络攻击能力,是智能网络攻防关键技术之一. 现有研究通过将系统审计日志建模成可表达攻击事件因果依赖关系的溯源图,利用溯源图强大的关联分析和语义表达能力,对复杂隐蔽网络攻击进行调查,相较传统方法效果提升显著. 在全面收集分析基于溯源图的攻击调查研究工作的基础上,根据溯源图利用方式及特征挖掘维度的差异,将基于溯源图的攻击调查方法划分为基于因果分析、基于深度表示学习和基于异常检测三类,总结凝练每类方法具体工作流程和通用框架. 梳理溯源图优化方法,剖析相关技术从理论向产业落地的能力演变历程. 归纳攻击调查常用数据集,对比分析基于溯源图的攻击调查代表性技术和性能指标,最后展望了该领域未来发展方向.

关键词: 攻击调查;溯源图;高级持续性威胁;深度学习;异常检测
基金项目: 国家重点研发计划(No.2022ZD0119602);国家自然科学基金(No.62272114);鹏城实验室重大项目(No.PCL2022A03);CCF-绿盟科技"鲲鹏"科研基金(No.CCF-NSFOCUS2023003)
中图分类号: TN915.08 **文献标识码:** A **文章编号:** 0372-2112(2024)07-2529-28
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20231057

A Survey of Network Attack Investigation Based on Provenance Graph

QIU Jing[1,2], CHEN Rong-rong[1], ZHU Hao-jin[3], XIAO Yan-jun[4], YIN Li-hua[1], TIAN Zhi-hong[1*]
(1. *Cyberspace Institute of Advanced Technology*, *Guangzhou University*, *Guangzhou*, *Guangdong* 510555, *China*;
2. *Pengcheng Laboratory*, *Shenzhen*, *Guangdong* 518000, *China*;
3. *Department of Computer Science and Engineering*, *Shanghai Jiao Tong University*, *Shanghai* 200240, *China*;
4. *NSFOCUS Technologies Group Co.*, *Ltd*, *Guangzhou*, *Beijing* 100089, *China*)

Abstract: Investigating network attacks is crucial for the implementation of proactive defenses and the formulation of tracing countermeasures. With the rise of sophisticated and stealthy network threats, the need to develop efficient and automated methods for investigations has become a pivotal aspect of advance intelligent network attack and defense capabilities. Existing studies have focused on modeling system audit logs into provenance graphs that represent causal dependencies of attack events. Leveraging the powerful associative analysis and semantic representation capabilities of provenance graphs, complex and stealthy network attacks can be effectively investigated, yielding superior results compared to conventional methods. This paper offers a systematic review of the literature on provenance-graph-based attack investigation, categorizing the diverse methodologies into three principal groups: causality analysis, deep representation learning, and anomaly detection. For each category, the paper succinctly presents the workflows and the core frameworks that underpin these methodologies. Additionally, it delves into the optimization techniques for provenance graphs and chronicles the evolution of these technologies from theoretical constructs to their application in industrial settings. This study methodically aggregates and reviews datasets prevalently utilized in attack investigation research, offering a comprehensive comparative analysis of representative techniques alongside their associated performance metrics, specifically within the ambit of provenance graph-based methodologies. Subsequently, it delineates the prospective directions for future research and development within this specialized field, thereby providing a structured roadmap for advancing the domain's academic and practical applications.

收稿日期:2023-11-15;修回日期:2024-05-29;责任编辑:宋甲英
*通讯作者:田志宏

Key words: attack investigation; provenance graph; advanced persistent threat; deep learning; anomaly detection
Foundation Item(s): National Key R&D Program of China (No.2022ZD0119602); National Natural Science Foundation of China (No.62272114); Major Key Project of PCL (No.PCL2022A03); CCF-NSFOCUS Kunpeng Funds (No.CCF-NSFOCUS202303)

1 研究背景

近年来,日益猖獗的网络攻击活动造成了重大经济损失.2010年“震网”病毒[1]的出现,标志着“网络战”元年的开启,网络空间安全威胁发生巨大变化.网络空间与现实空间紧密关联,但又平行独立,科技进步助力网络攻击、新技术的新安全问题等不同维度因素,也在深入影响全球网络空间安全秩序.随着我国经济快速发展和国际地位的不断攀升,我国已成为全球APT(Advanced Persistent Threat)[2~5]活动首要地区性目标之一.现代网络攻击具有多步性[6,7]和隐蔽性[8,9].多步性是指网络攻击行为分多个步骤、跨越多台主机,手段复杂,受害面积广;隐蔽性主要来自攻击链中未知漏洞利用、安全设备绕过与加密流量使用等.攻击调查[10]是应对上述难题的重要手段,以少量警报点为线索,进行攻击事件前后向关联分析,以还原完整攻击路径.传统网络攻击调查主要依靠安全分析人员手动关联攻击场景,通过日志检索,结合专家知识,发现攻击入口点,确认攻击受害面.然而现代网络攻击面积覆盖大、时间范围广、手段隐蔽性强,端点和流量监控设备记录的攻击日志条目之间存在跨度大、间隔长、难检索、难关联等问题,传统的以原始审计日志为对象的攻击调查方法,由于缺少事件间直接的因果关联表达,精准完整的攻击场景调查困难重重.

针对上述挑战,越来越多研究者们致力于将系统日志建模成可表达事件信息流的溯源图[11],旨在通过这种语义丰富的结构化数据表示提高攻击调查任务性能.溯源图是系统实体和系统事件的集合,系统实体表示为节点,系统事件表示为边,两个节点之间可以存在多条时间不同或类型不同的边.一个Web攻击事件对应的溯源图如图1所示,展示了攻击者利用Apache服务漏洞获取权限,从恶意服务器上下载后门程序并执行的过程.

溯源图将审计日志中的实体和关系抽取出来,能够将因果相关的攻击事件直接关联,大大提升攻击调查效率与精度.溯源图的抽象层次合理,由系统实体和系统事件组成的溯源图能够在主机层面记录全量攻击行为,保证了攻击调查的完整性;溯源图语义信息丰富,包含时空属性和上下文语义等多个维度,基于多维语义信息,引入先进的智能算法,能够将溯源图中的行为提取出来,通过学习攻击模式发现已知攻击,或通过学习正常行为发现异常行为,进而调查隐蔽未知攻击

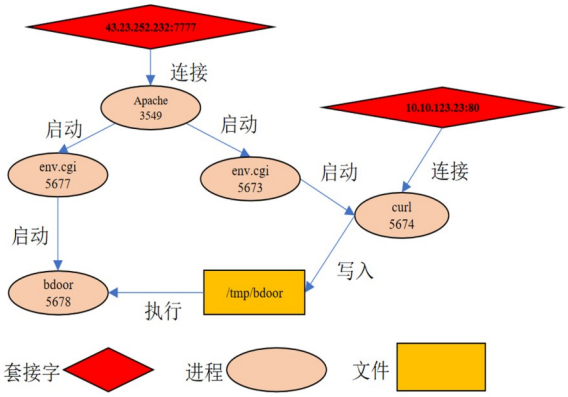

图1 Web攻击溯源图实例

线索,实现精准完整攻击调查.我们以“溯源图”“攻击调查”为关键词,在Scopus和WoS数据库中进行了检索,筛选了近5年的研究成果,构建的频繁词知识图谱如图2所示.

近几年虽已存在溯源图相关研究的理论及应用工作梳理[12~18],但针对基于溯源图的攻击调查方法缺少完整的体系化总结.本文贡献主要体现在四方面:

(1)针对基于溯源图的攻击调查问题本身进行剖析,提炼核心需求;

(2)依据攻击语义特征表达,对基于溯源图的攻击调查方法进行分类总结,提出了更接近分析人员高层语义理解的全新分类思路;

(3)通过对已有研究成果进行梳理,针对上述分类方法,总结出各方法实现攻击调查任务的流程框架及关键技术,为领域研究工作提供参考和指引;

(4)基于前沿成果和发展形势,本文提出了该领域发展趋势、面临的问题与挑战及未来研究方向.

2 基于溯源图的攻击调查方法及相关技术

通过引入溯源图,相关的攻击事件能够在系统实体粒度上直接关联,从根本上优化数据表示方法,提升攻击调查效率.近年来,基于溯源图的攻击调查方法成为学界和企业的研究热点,成为主流的攻击调查方法,相关研究主要集中在以下两个需求.

(1)面向溯源图的攻击语义特征挖掘利用

现代网络攻击场景复杂,手段隐蔽.待调查溯源图中存在许多攻击无关节点和关系干扰,如何挖掘溯源图中的显式语义特征,指导因果分析优先搜索攻击相

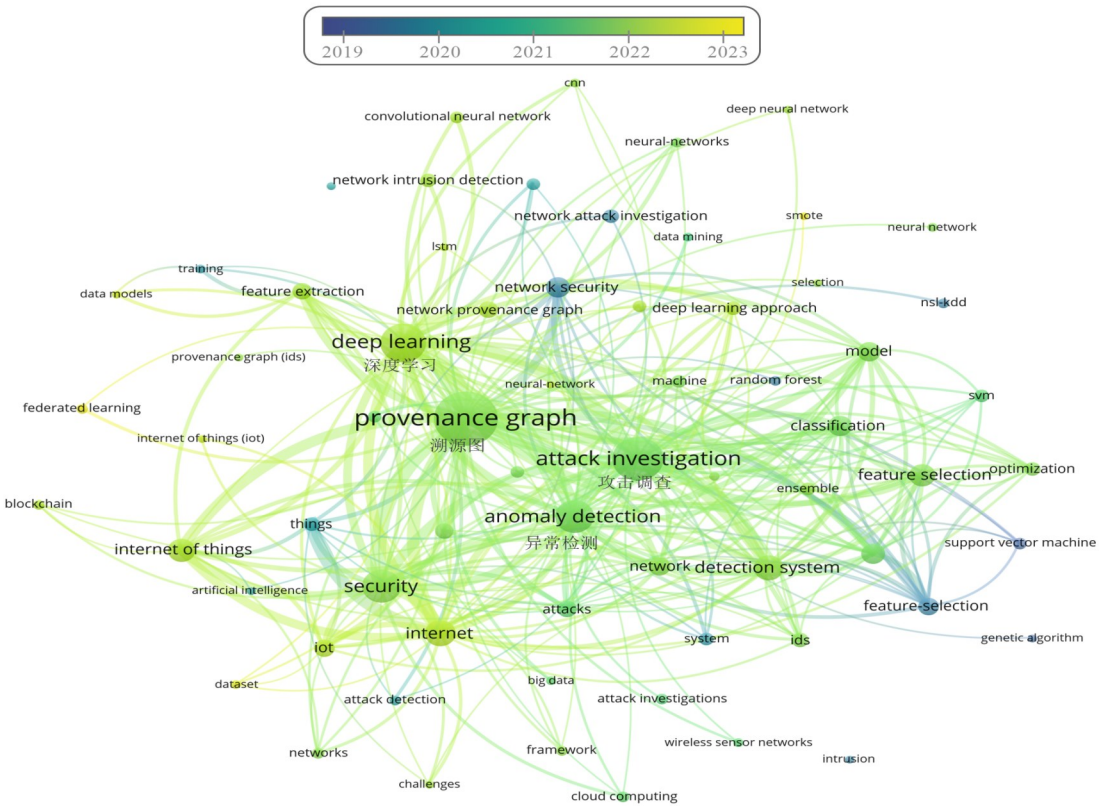

图2　基于溯源图的攻击调查研究热点关键词知识图谱

关事件,基于深度隐式特征发现更多攻击线索是需要研究的一个重要内容.

(2)面向溯源图的结构属性优化压缩

由于现代软件系统的复杂性和网络攻击的长期性,溯源图规模庞大,导致存储压力大、计算复杂度高.这是溯源图实现产业化落地的痛点,因此如何在不丢失攻击信息的前提下,对溯源图结构和属性进行压缩优化是亟待解决的问题之一.

2.1　面向溯源图的攻击语义特征挖掘利用

本文根据对溯源图语义的挖掘程度以及技术发展趋势,将基于溯源图的攻击调查方法分为三类:基于因果分析的攻击调查方法、基于深度表示学习的攻击调查方法和基于异常检测的攻击调查方法.随着网络安全环境的变化和先进智能算法的发展,基于溯源图的攻击调查技术持续迭代,研究重点也不断发生转变.

纵观技术发展历史,早期相关研究主要基于溯源图上的直接因果分析,根据已知攻击线索,在图上进行回溯与前向分析,找到攻击入口点,发现攻击的完整影响,完成攻击路径或场景搜索任务.但由于溯源图中大量无关节点干扰,直接因果分析调查效率差、精度低,近些年相关研究开始在图上标注节点标签、路径权重等显式辅助语义信息,指导深度优先搜索、最短路径等各类图算法优先搜索攻击相关事件,加速因果分析.这类方法利用了溯源图的显式语义,但是在进行图上搜索的时候主要依赖预先定义的专家规则,且获取到的上下文信息不够完整,缺乏对溯源图上节点行为的深刻理解,仅利用规则挖掘显式语义特征,发现攻击线索能力有限.

通过引入深度表示学习智能算法,能够补充因果分析方法中缺失的上下文隐式语义,用高维向量表示溯源图上节点行为,提供更多攻击线索.近年来,图神经网络等深度学习算法给溯源图领域研究者带来新启发,通过提取溯源图中的攻击序列或攻击行为子图,提取溯源图的隐式语义特征,输入到深度学习模型中,学习包含上下文信息的攻击行为模式,能够高精度检测攻击,提供攻击线索.然而有监督深度学习算法需要大量标注训练数据,目前领域缺乏高质量的标注数据集支撑优质模型训练.且攻击检测与调查任务中获取到的日志数据具有显著的类别不平衡问题,良性数据比例通常在99%以上,简单的数据增强方法难以提升模型的泛化能力.此外,基于训练数据的模型难以应对未知攻击,上述问题限制了基于深度表示学习方法的发展.

异常检测方法利用显式隐式融合特征,为攻击线索发现提供了新思路.异常检测是指在不依赖标注数据的前提下学习正常行为模式,发现异常攻击行为.隐

蔽未知攻击能够逃避基于规则和先验知识的检测系统,却很难完全伪装成正常行为,因此异常检测系统能够补全大量的攻击线索,支撑攻击场景重建与调查.基于异常检测的方法融合了因果分析和深度表示学习的优势,结合显式语义特征和隐式上下文语义特征,将溯源图转化为可度量对象,通过阈值判断或无监督聚类等方法发现离群异常点,发现隐蔽攻击行为.然而异常检测方法也存在一定局限性,包括较高的误报率以及随之而来的警报疲劳等问题.因此,如何加深对攻击行为的理解,提取更有效的融合语义特征,提升异常检测方法的准确率,降低误报率是未来的研究重点.

基于以上分析,本文梳理出三类基于溯源图的攻击调查方法基本框架.

(1)基于因果分析的攻击调查方法

基于因果分析的方法主要挖掘溯源图的显式语义特征,利用溯源图中表达的事件因果关系,以及警报事件特征,对溯源图进行增强,标注攻击语义,结合上下文因果分析,对攻击场景进行重建调查.

(2)基于深度表示学习的攻击调查方法

基于深度表示学习的方法侧重挖掘溯源图隐式语义特征,通过引入先进智能算法,学习溯源图的拓扑结构语义、时空语义与攻击行为语义等特征,发现攻击模式,辅助攻击调查任务.

(3)基于异常检测的攻击调查方法

基于异常检测的方法综合分析溯源图显式特征和隐式特征,将溯源图转化为可度量对象,基于统计学习方法量化异常度,进而发现异常,提供更多攻击线索供攻击调查使用.

图3是基于溯源图的攻击调查方法细分框架,展示了三类方法的细分类别和对应代表性文献.基于因果分析的攻击调查方法面临着实时调查困难的问题,研究者们提出多种方法加速图上因果分析计算,根据语义分析方法的不同,将基于因果分析的攻击调查方法分为基于语义添加和基于语义分区的方法.基于深度表示学习的攻击调查方法旨在从不同维度挖掘模式,加深对攻击和系统进程、用户行为的理解,辅助攻击调查,根据模式分析对象的不同,将基于深度表示学习的攻击调查方法分为基于攻击模式、基于进程模式和基于行为模式的方法.基于异常检测的攻击调查方法早期多以子图作为检测对象,以子图表征系统状态,分析粒度较粗,为了精确感知定位威胁,更多的研究聚焦到路径级的异常检测,发现异常路径和节点,根据检测粒度的不同,将基于异常检测的攻击调查方法分为基于子图分析和基于路径分析的方法.

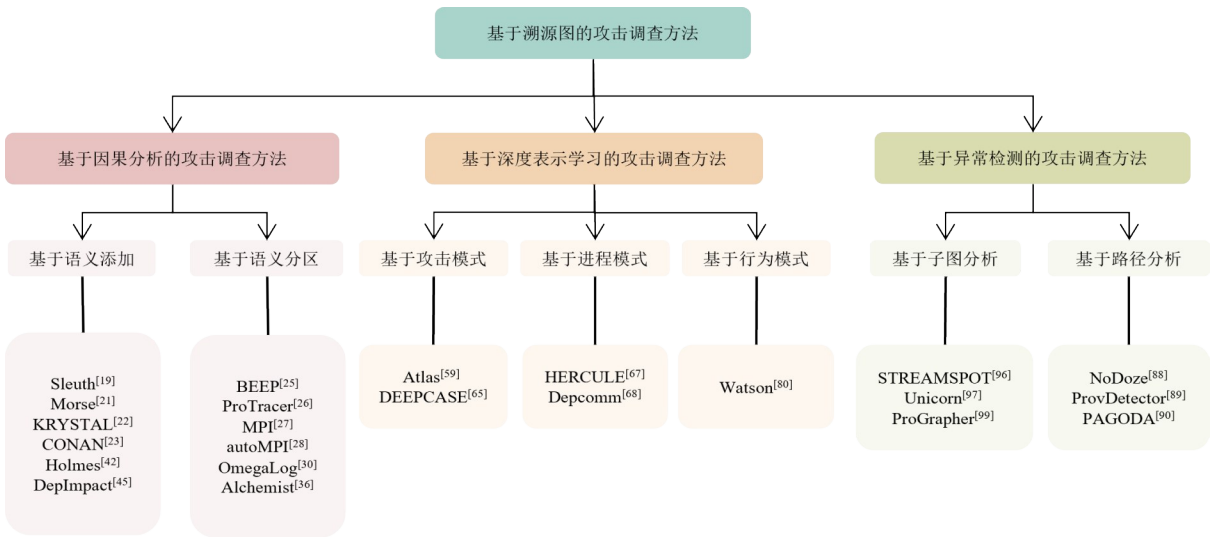

图3 基于溯源图的攻击调查方法分类细分框架

2.2 面向溯源图的结构属性优化压缩

本文全面分析溯源图结构属性优化问题,现有相关研究可以分为基于有损压缩和基于无损压缩的方法,有损压缩方法主要目标是减小图规模,对冗余数据进行剪枝;无损压缩方法通过更好的编码(即使用更少的位存储相同的数据)来缓解存储压力.有损压缩方法能够剪除攻击无关的边,在减小图规模的同时提高调查效率,但剪除的“冗余”数据中可能包含攻击信息,影响攻击调查完整性.无损压缩方法保留全量信息,但是带来了压缩和解压缩的额外开销,全量数据中的攻击无关信息给攻击调查带来阻碍.因此平衡调整有损压缩和无损压缩方法的使用范围,达到攻击调查效率和完整性的平衡是溯源图优化技术发展的关键问题.压缩溯源图存储和计算是搭建基于溯源图的攻击调查系统的基石,优化压缩溯源图规模是产业落地的必由之路.

2.3　相关综述工作

目前已有一些文献[12~18]对溯源图相关工作进行了整理分析. 文献[13~16]对已有的溯源调查技术进行分类,其分类依据是具体技术方案,以算法为线索,将攻击调查区分为基于启发式算法的方法、基于语义相似性的方法和基于机器学习的方法等. 文献[12]主要从图数据优化和上下文分析两个方面探讨了攻击溯源技术的现状和发展方向. 对于图数据优化问题,文献[12]介绍了细粒度数据收集和数据完整性保证的技术方案,讨论了不同方法的优势和不足,指出它们在实际应用中面临的挑战. 上述文献偏重于算法、技术的分类,但在溯源调查技术的框架性概括总结上稍显不足. 文献[13,14]梳理出完善的基于溯源图的威胁检测系统架构,将系统分为数据采集、数据管理、威胁分析与调查等模块. 数据采集模块包含各个粒度,不同层级的审计日志收集场景;数据管理模块主要介绍了溯源图剪枝、溯源图存储与查询的常见方法;威胁分析与调查模块提出了在线的威胁狩猎,离线的威胁检测,基于警报完成取证调查任务,从工作流角度提出溯源调查系统的一般框架,但缺乏流程内部细粒度的机制分析. 文献[15]将基于溯源图的攻击调查分为五个层次:采集、优化、查询、检测和调查,同时针对图优化层问题开展实验分析. 为基于溯源图的攻击调查研究应用落地提供系统性指导,该工作叙述重心在溯源图构建与优化层面,攻击检测与调查的相关论述较少. 相较上述工作,本文侧重于对研究思路的发展概括分析. 从溯源图的语义挖掘程度出发,对溯源图的发展趋势、面临的问题挑战和应用场景等方面进行总结与归纳. 梳理基于溯源图的攻击调查方法技术路线及能力演变历程,旨在为研究人员提供先进思路,迅速定位研究语义层级,为基于溯源图的攻击调查研究和应用落地提供参考.

3　基于因果分析的攻击调查方法

基于因果分析的攻击调查方法致力于高效、快速重建精准完整的攻击场景,但溯源图的庞大规模对调查效率提出了挑战. 溯源图以系统实体为节点,增长十分迅速,冗余信息多,威胁行为隐藏在海量数据当中,深度或广度优先路径搜索计算复杂度高. 研究人员尝试通过语义添加来加速路径搜索,前期工作通过为节点添加威胁标签信息,定位攻击行为,并基于专家经验和标签信息为边赋予权重,根据标签和权重,重建攻击场景. 基于标签的方法对专家经验和预设规则要求高,后续工作尝试从溯源图本身挖掘语义,减少对专家知识的依赖,从图结构、警报、时间等多个维度为边赋予权重,指导路径搜索算法优先分析高置信威胁场景.

另一方面,随着计算机系统的长时间运行,持续运行的进程,如浏览器进程、Web服务进程等会产生大量的边,这些边代表时间段内的多个行为,并非每个行为都与攻击相关,导致路径搜索算法遍历大量无关节点和边,影响攻击调查和安全人员研判分析效率. 研究人员提出语义分区方法解决这个问题,相关工作首先着眼于程序内部运行逻辑,通过分析程序二进制层面控制流,或利用插桩技术输出调试信息,划分溯源图行为分区. 但该类方法只适用于开源软件,且二进制分析难度高,插桩开销大,应用范围受限. 最新的工作尝试利用应用程序日志,将每次行为单独分区,减少无关依赖影响,确保攻击调查速度和准确性. 因此基于因果分析的攻击调查方法可分为基于语义添加和基于语义分区的方法,图4展示了基于因果分析的攻击调查方法的分类和代表性文献.

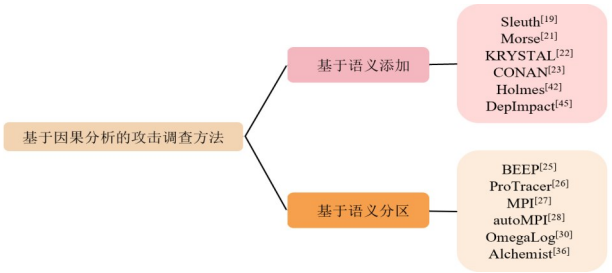

图4　基于因果分析的攻击调查方法分类图

基于因果分析的攻击调查方法,能够直接利用溯源图中表达的事件或数据因果关系进行上下文分析,结合添加的语义增强信息或拆分的语义分区,以警报事件为基点,执行回溯关联分析,重建完整的攻击场景. 如图5所示,基于因果分析的攻击调查可分为三个步骤:溯源图增强、攻击威胁评估和攻击调查.

3.1　溯源图增强方法

原始溯源图规模庞大,结构冗余,语义信息不显著,通过添加辅助信息,能够对溯源图进行语义升级,结构剪枝,形成增强溯源图,支撑后续高效因果分析与攻击调查. 根据优化侧重点不同,溯源图增强方法可分为基于辅助语义的增强方法和基于语义分区的增强方法.

3.1.1　基于辅助语义的增强方法

为溯源图添加标签辅助语义,利用标签在溯源图上表达与传播攻击相关语义,增强溯源图. 过程包括标签初始化、标签传播和标签衰减等几个阶段. SLEUTH[19]根据攻击必由阶段(如恶意代码部署与执行)和攻击常见目的(如数据窃取)设计了一套标签规则. 该系统中包含两种基础标签类型,分别是代表节点受信任程度的可信度标签T-Tag(Trustworthiness Tag)以及代表节点数据机密性的机密性标签C-Tag(Confiden-

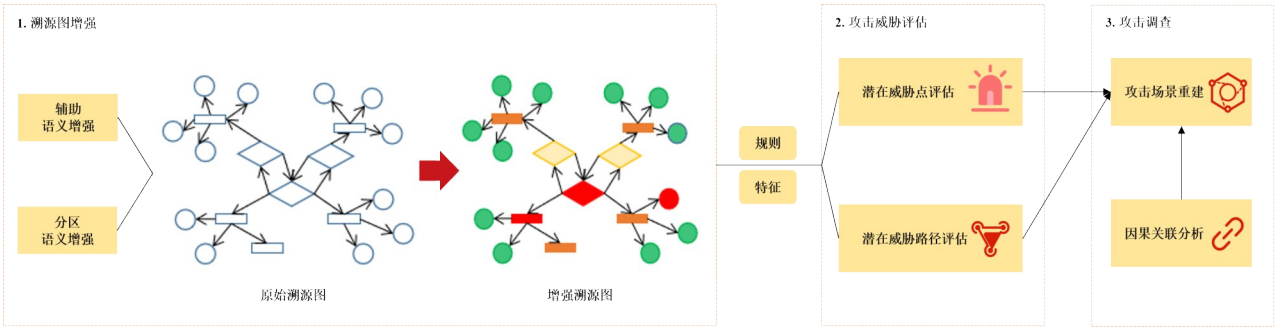

图5 基于因果分析的攻击调查方法框架图

tiality Tag). 针对这两类标签,SLEUTH设计了标签初始化和传播规则,标签初始化的主要目的是为溯源图中的某些特殊语义节点赋予默认标签,例如对敏感文件赋予高机密性标签,将内网网络连接节点初始化为高可信标签,来自外网的网络连接初始化为低可信标签等. 标签传播[20]是为了将初始化标签语义传递到整张图上,根据预先定义的事件传播规则,为整张溯源图赋予标签,提高溯源图的语义丰富程度. 图6是基于标签规则的溯源图增强方法流程示意图,由于可信度标签与机密性标签过于复杂,不便展示,故仅在图中展示危险标签、可疑标签、未知标签和良性标签.

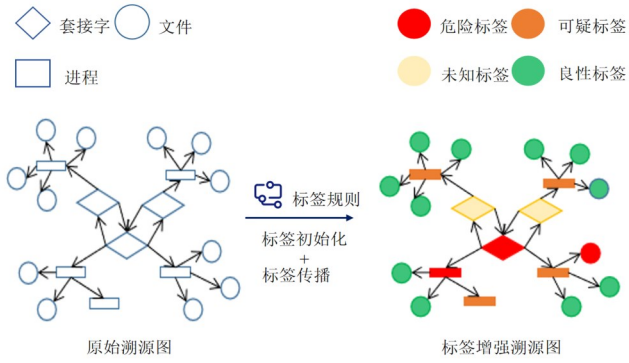

图6 基于标签规则的溯源图增强方法流程图

但由于标签传播规则的局限性,某些长时间运行的关键节点如果被标签污染,可能会将恶意标签传播到整张图上,MORSE[21]和KRYSTAL[22]通过引入标签衰减策略来缓解标签污染问题. 它们将标签从类别转化为数值,细化了标签语义粒度,继而引入衰减策略对标签数值进行衰减. 如果某个节点接收了可疑输入,那么攻击行为应该紧随其后,随着时间推移,可疑输入带来的影响应该周期性衰减(对应攻击者的行为规律),所以标签衰减策略的原则是:节点的威胁程度随时间周期性变化. CONAN[23]和APTSHIELD[24]对标签语义进行扩充,补充了进程的行为和状态语义,完善了标签规则,支撑更加复杂语义和外部知识的引入. 但是对标签语义进行扩充,引入更复杂的行为和状态语义,增加了标签规则的复杂性,这可能使得方法更难以理解、配置和维护. 标签初始化和传播的过程中需要一定的计算资源,在大规模和实时的网络环境中,计算复杂性可能成为一个挑战.

3.1.2 基于语义分区的增强方法

对溯源图进行语义分区,获取细粒度因果关系,可以避免依赖爆炸问题. 在利用溯源图进行攻击调查时,某些长时间运行的进程会产生大量依赖,如图7所示,火狐浏览器进程可能连接着许多网络IP地址和端口,但实际上浏览器进程的行为起因可能是其中一个网络连接,此时进行图上因果分析的计算消耗将大幅增加,称为依赖爆炸. 为了解决依赖爆炸问题,研究者们提出语义分区的方法,拆分进程行为的细粒度因果关系,能够剔除与攻击无关的事件,使得溯源图上攻击调查分析更加高效.

BEEP[25]和ProTracer[26]基于逆向工程方法,对应用程序二进制文件中的循环以及迭代依赖关系进行识别,将长时间运行进程划分为自治的执行单元. 通过这种方式,输出事件只依赖于同一执行单元中前面的输入事件. MPI[27]和autoMPI[28]采用插桩技术[29],在保证原有程序逻辑完整性基础上,向程序中插入探针,通过探针采集代码中的信息(方法、参数、返回值等),在特定的位置插入代码段,收集程序运行时的动态上下文信息,分析程序执行单元,获取精准的事件因果依赖关系. OmegaLog[30]是一种利用应用程序日志进行语义拆分的方法,基于对程序二进制文件的静态分析[31,32]和符号执行[33~35],获取程序写入的应用日志消息字符串,并执行时序分析,获取所有可能的应用日志消息序列. 通过这些序列,将应用程序行为划分为执行单元,解析审计日志流. 但OmegaLog无法区分同一工作线程中不同单元的子任务,当异步行为密集时,它无法保证因果分析的准确度. 上述几类方法都需要对源代码进行分析,获取程序隐私数据,真实环境应用价值较小. 相比之下,Alchemist[36]使用datalog[37]引擎推断出丰富的语义信息,例如并发任务中交错的原子部分,以及应用程序日志或审计日志中不可见的依赖关系,将审计日志

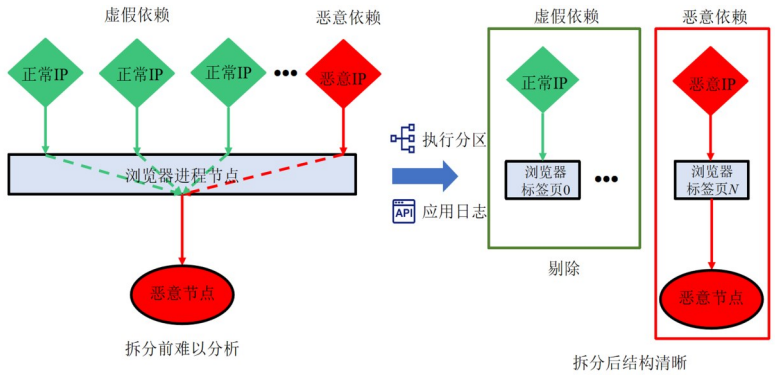

图7　溯源图语义分区方法示意图

与应用程序日志融合. 一方面,丰富的应用程序语义可以传播到系统调用级别,另一方面,审计日志中记录的低层级背景信息可以准确地分区到高级应用程序执行单元,从而排除虚假的依赖关系. 但 Alchemist 需要对主机上每个应用程序设计日志解析器,工作量庞大. MCI[38]基于 LDX 因果模型[39]对溯源图事件序列进行建模,从而推断系统调用之间的精确因果关系,对复杂依赖的节点进行分区. 但因果模型的计算开销较大,应用存在一定局限性. UISCOPE[40]通过引入用户界面(User Interface,UI)[41]事件对应用程序行为进行分区,其核心思想是对 UI 元素/事件和低级系统事件进行因果分析,将系统事件与 UI 事件相关联,长时间运行的进程被划分为单独的 UI 转换,填补了低级系统事件与 UI 事件之间的语义鸿沟,使因果分析结果准确,一定程度上解决依赖爆炸问题,支撑高性能攻击调查. 上述方法在解决依赖爆炸问题时提供了一些创新的思路,但应用条件都有所限制,包括源代码的可获取性、因果模型的实时性、大规模部署的可行性等,所以如何有机地统合各类方法,形成适用范围广、可部署性强的解决方案十分关键. 且所有方法都面临语义分区的精确性问题,在实际应用中,如何确保语义分区的结果对攻击调查有意义,以及如何避免虚假的依赖关系,是一个需要深入研究的问题.

3.2　攻击威胁评估方法

攻击调查是基于攻击线索开展的,根据攻击线索可以对溯源图进行攻击威胁评估. 攻击威胁评估可以区分为两个层次:潜在威胁点评估和潜在威胁路径评估,潜在威胁点评估主要是发现更多的攻击线索,点亮攻击场景中的关键节点;潜在威胁路径评估的目的是表示因果事件与实际攻击行为之间的关联程度,提高因果分析效率.

3.2.1　潜在威胁点评估

潜在威胁点评估旨在发现更多攻击线索,在图上划定警报与威胁范围,作为攻击调查依据. SLEUTH 等[19,21]根据增强溯源图上的标签,制定警报规则,如敏感数据泄露,在一次进程对外发送网络连接事件中,如果进程的可信度标签是未知,机密性标签为敏感,将会触发敏感数据泄露警报,从而发现攻击事件. 基于标签规则的威胁点评估速度快,规则具有可扩展性,但其对专家知识的高度依赖使得应用场景受限. 通过引入外部标准知识库可以缓解现有警报策略对专家知识的依赖问题,ATT&CK 框架是一种综合性的敌方战术、技术和程序(TTPs)知识库,用于分析和分类网络威胁. TTPs 是攻击者用来针对目标实施攻击的具体方法技术. HOLMES[42]将新生成的溯源图流式输入 TTP 报警引擎,与预定义攻击行为规则进行匹配,产生报警,图 8 是几个警报的示例. KRYSTAL[22]将溯源图构建成知识图谱[43],通过引入开源 SIGMA 规则,将警报语义标注任务转化为知识图谱上的 SPARQL[44]查询任务,评估溯源图上的警报语义. 随着攻击手法和威胁情境不断演化,标签规则和知识库需要及时更新. 如何实现标签规则和知识库的动态演化,以适应新的攻击方式至关重要. 在面对对抗性攻击时,攻击者可能会有意避开已知的规则和模型,上述方法缺乏对抗性,难以应对未知攻击. 因此如何提高检测方法对抗性,以未知对抗未知,是亟须研究的问题.

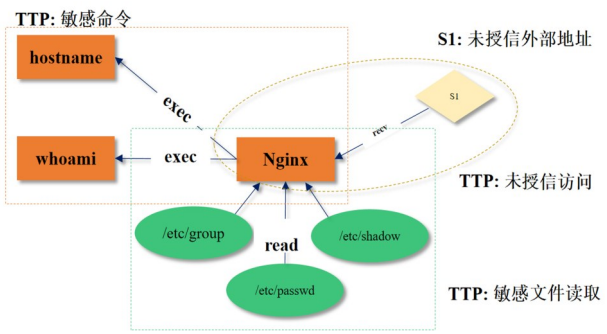

图8　基于 TTP 规则匹配的警报示例

3.2.2　潜在威胁路径评估

通过对所有潜在威胁点形成的威胁路径进行评

估,指导进行攻击因果分析.潜在威胁路径评估是以溯源图上标注的增强信息和警报威胁作为启发式信息,利用溯源图中因果依赖关系,传递或传播攻击语义,实现溯源图上边或节点的权重赋值,帮助完整攻击场景重构.SLEUTH[19]设计了一种基于标签的边权值计算方法,在标签溯源图上,根据专家知识对攻击规律进行总结提炼,形成赋权规则,与攻击相关程度越高的事件权值越小,例如低可信度主体执行良性客体,将这条边权值赋为0,因为该事件极可能与攻击相关.基于标签的权重评估较快,计算量较小,但是缺乏对溯源图语义特征的理解.DEPIMPACT[45]挖掘更多维度的溯源图语义特征,判断事件与攻击的相关程度,对溯源图的边进行权重赋值.特征主要包含三种:时间特征、图结构特征和流量特征.时间特征反映了事件与警报之间的时间间隔,与警报时间越接近,越有可能是攻击相关事件.图结构特征挖掘了图上节点与边的聚集关系,若从警报事件出发,通过多条路径都可以回溯到同一个事件,则该事件很可能是攻击的一个关键步骤.流量特征表达了事件与攻击者部署载荷之间的相关程度,若图中两点之间的数据流量与警报事件(若为载荷部署类警报,如Webshell[46]、远程控制木马[47,48]等)的流量有相似的大小,则认为该事件与警报事件相关性更高.通过使用基于kmeans++[49]和LDA[50]算法的区别特征投影方法,将三种特征进行加权计算,得到每条边的最终权重.DEPIMPACT结合多维特征对溯源图上的边进行评分赋权,准确性更高,但是需要收集和处理大量的数据,计算成本较高.HOLMES通过引入CVSS(公共漏洞评分系统)[51]知识库,CVSS将TTP的威胁等级分为低、中、高、严重,每个等级分配不同的分数,对每个匹配到的TTP警报赋予权值.然而,CVSS的威胁等级可能无法全面覆盖所有攻击场景,对于新兴威胁可能不够敏感.在上述方法中,权重的赋值主要基于专家知识、攻击规律、特征挖掘等.然而,由于攻击的多样性和复杂性,很难确保权重的准确性和有效性.尤其是当面临新的攻击手段或复杂的攻击场景时,现有方法可能无法准确评估威胁路径,如何设计一种通用的、可扩展的威胁评估方法是未来的研究方向.

3.3 基于因果关联分析的攻击调查方法

基于增强溯源图和评估的攻击威胁,能够执行前后向因果关联分析,将攻击语义关联起来,搜索攻击入口点,发现攻击的完整影响,自动化重建精准、完整的攻击场景.基于标签的方法通过回溯因果分析找到攻击的入口点,并根据边上的权重,计算从警报点到入口点的最短路径,发现攻击的真正入口点.然后从入口点执行前向分析,通过深度优先搜索算法,根据路径长度阈值限制,还原出一幅完整的攻击场景图.如图9所示,HOLMES[42]将TTP警报进行关联,形成高级场景图,基于七阶段杀伤链模型[52],将高级场景图抽象表示为七元组,并进行数值化.每个阶段可能含有多个TTP,取其中威胁程度最高的TTP代表这个阶段,将其分数赋到七元组中,进行加权乘积计算,得到高级场景图的总评分,最后进行阈值比较,如果高于阈值则认为该高级场景图是真正的攻击,否则是虚假的攻击,从而筛除虚假警报.DEPIMPACT[45]受TrustRank算法[53]启发,将分数传播思想应用到攻击调查,以警报节点为起始点,将其权值赋为1,根据边上的权重将影响传播到入口点,从分数较高的入口点开始正向分析,即可找到攻击路径和潜在影响.在网络攻防中,快速响应至关重要.采用深度优先搜索、反向传播进行前后向分析时,随着图规模增大,计算耗时可能显著增加,影响及时发现和处置安全事件的能力.所以如何减少图规模增长对攻击调查的影响,提高计算效率是一个挑战.

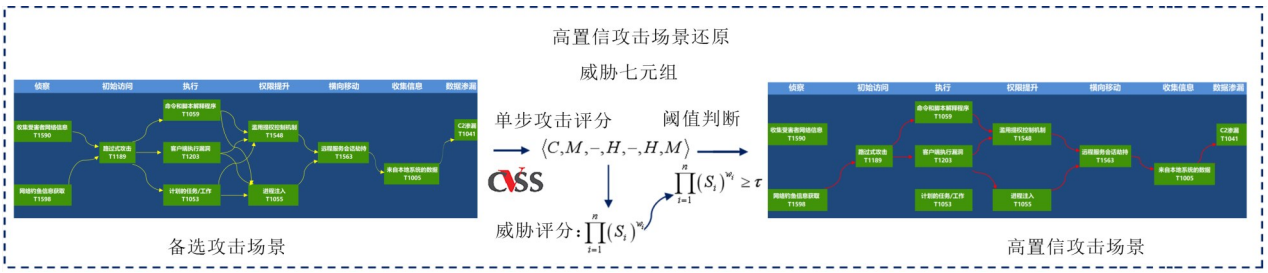

图9 基于威胁建模评分的高置信攻击场景还原方法

3.4 小结

基于因果分析的攻击调查方法挖掘了溯源图的图结构特征,通过标签语义、外部知识等高层语义信息发现警报事件与攻击行为之间的因果关联关系,解析攻击者的技战术手段,核心目标是高效率重建精准完整的攻击场景.找得快、找得准主要依赖对溯源图上节点和边的威胁语义量化评估,优先搜索高威胁的节点和边,因此基于攻击理解挖掘溯源图中攻击相关语义信息十分重要;但标注语义信息的同时需要兼顾算法效率,所以语义信息的粒度也是一个关键问题.但基于因果分析的调查方法对专家或领域知识规则依赖程度高,对溯源图中因果依赖分析层次浅,随着先进智能算

法的发展,越来越多的研究者尝试引入智能算法进行攻击路径权重计算,或基于溯源图拓扑结构学习溯源图中的隐含深度特征.

4　基于深度表示学习的攻击调查方法

基于深度表示学习的攻击调查方法能够挖掘各类隐式特征,构建模型发现攻击. 溯源图含有丰富的上下文语义信息,研究人员首先关注攻击模式在溯源图上的表达,利用少量带标注数据构造攻击序列和非攻击序列训练模型,然而攻击相关数据的稀有性限制了攻击模式学习的发展. 进程作为系统中发起事件的主体,是溯源图中最重要的部分,研究人员设计元路径进行随机游走,结合社区划分算法自动学习进程模式,形成以进程为核心的社区,结合攻击线索还原攻击场景. 进程处于系统实体粒度,形成的模式与安全人员理解仍旧存在语义鸿沟. 攻击者可能改变其使用的攻击手段或工具,但上升到行为层面仍是类似的,因此提取行为模式能够填补语义鸿沟,帮助安全人员更好地理解攻击. 因此基于深度表示学习的攻击调查方法可分为基于攻击模式、基于进程模式和基于行为模式的方法. 图10展示了基于深度表示学习的攻击调查方法的具体分类和代表性文献.

基于深度表示学习的攻击调查方法通过在溯源图中聚合实体的上下文信息,获取溯源图中的隐式特征. 首先将溯源图拆分成不同粒度的攻击语义单元,获取多层次攻击相关信息,利用深度神经网络[54]学习对应攻击语义单元表征. 根据学习到的攻击模式和行为,辅助完成攻击调查. 图11是梳理出的本类方法框架图.

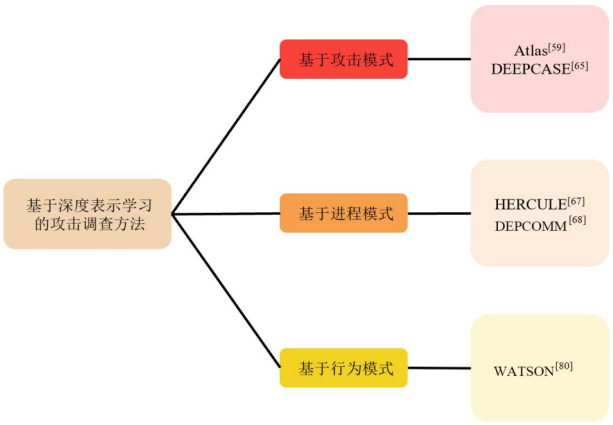

图10　基于深度表示学习的攻击调查方法分类图

4.1　攻击语义单元表征方法

4.1.1　基于序列的表征方法

序列[55,56]可以表示攻击模式[57,58],溯源图是系统活动的抽象表达,图中的序列可以表示连续的一串活动. 通过对攻击的观察发现,尽管攻击的实现方式可能不同,比如利用不同的漏洞或者执行不同的攻击载荷,但是复杂攻击可能具有相似的攻击模式. 而攻击在系统中又以事件的形式进行描述,因此可以通过学习具有时序的攻击序列,来发现攻击中包含的抽象攻击阶段之间的关系,提取攻击模式. ATLAS[59]从溯源图中的攻击实体和非攻击实体中,提取出攻击序列和非攻击序列进行学习. 首先获取一组攻击实体,构建出两个或多个攻击实体的子集,对于子集中的每个实体,提取它和其邻居对应的边作为攻击事件,子集实体涉及的所有攻击事件构成攻击序列. 由于非攻击实体的数量比较多,为了学习攻击序列和非攻击序列之间的差别,向每个攻击实体子集中添加一个非攻击实体,从而提取非攻击序列. 基于过采样[60,61]和欠采样[62,63]方法平衡攻击序列和非攻击序列的数量,利用LSTM模型[64]学习攻击的模式,图12是ATLAS对溯源图进行序列表征的流程示意图. 但采用过采样或欠采

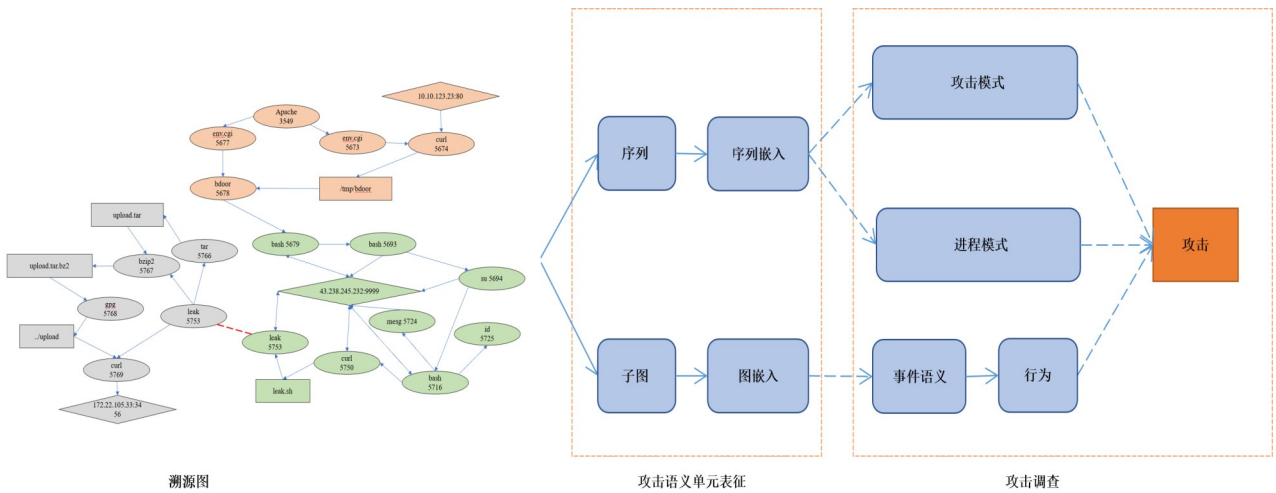

图11　基于深度表示学习的攻击调查方法框架图

样策略时，应当谨慎处理以避免过拟合或欠拟合现象，同时确保模型在未知数据上的泛化能力．DEEPCASE[65]是一种基于半监督方法[66]的安全事件上下文分析系统，针对安全事件序列进行嵌入表示，将相关事件进行聚类并呈现给安全人员，大大减少了分析规模．安全人员对事件序列模式进行标注，设置策略，基于反馈的策略可以忽略掉非关键事件的序列．

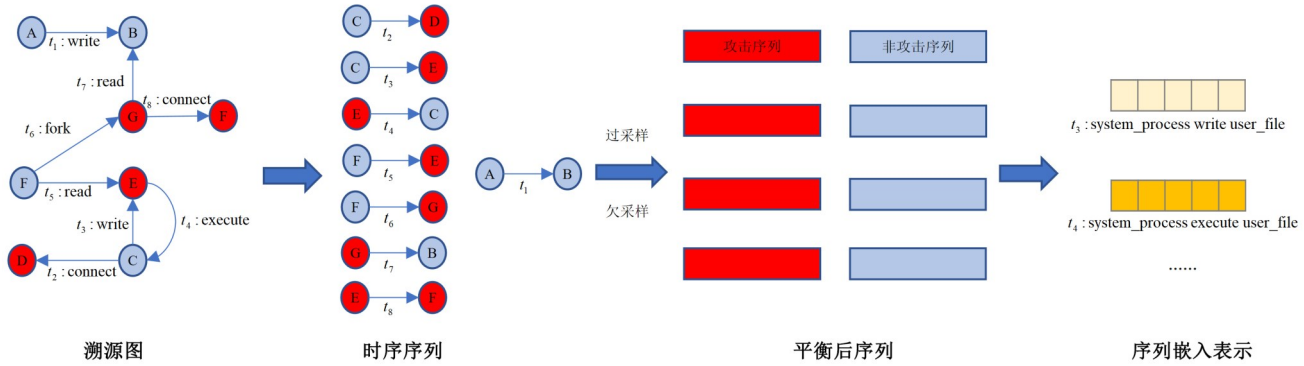

图12 基于邻域分解的溯源图序列采样方法流程图

序列还可以表示进程模式，在溯源图中，进程作为系统活动的发起者，是攻击行为的核心点，从溯源图中提取以进程为中心的社区来揭示进程与实体之间的因果关系，可以辅助攻击调查．HERCULE[67]将多个轻量级日志条目关联起来，构建多维加权图，从这些图中发现攻击社区．DEPCOMM[68]根据专家知识，设计了一套随机游走[69]策略，从进程节点出发，生成特定长度的序列，关系比较亲密的进程将会存在于一个序列中，基于Skip-gram算法[70]对节点进行学习，获取进程节点的嵌入表示．对进程节点进行聚类[71]，并根据聚类结果对资源实体进行分类并产生社区[72,73]．其中每个社区都包括一组亲密进程和这些进程访问的系统资源，表达了进程的模式．图13列举了随机游走的策略，描述了一个随机游走路径的具体实例．但该方法依赖于预设的随机游走策略和专家经验来定义节点间的联系强度，这种依赖可能会限制系统的自动化程度和面对新型、未知攻击时的反应能力．构建图、提取序列、训练模型以及生成社区结构的过程可能会消耗大量计算资源和时间，这与网络安全对即时性响应的要求相冲突．因此，如何构建轻量化模型，在检测效率和响应时间之间达到平衡是关键所在．

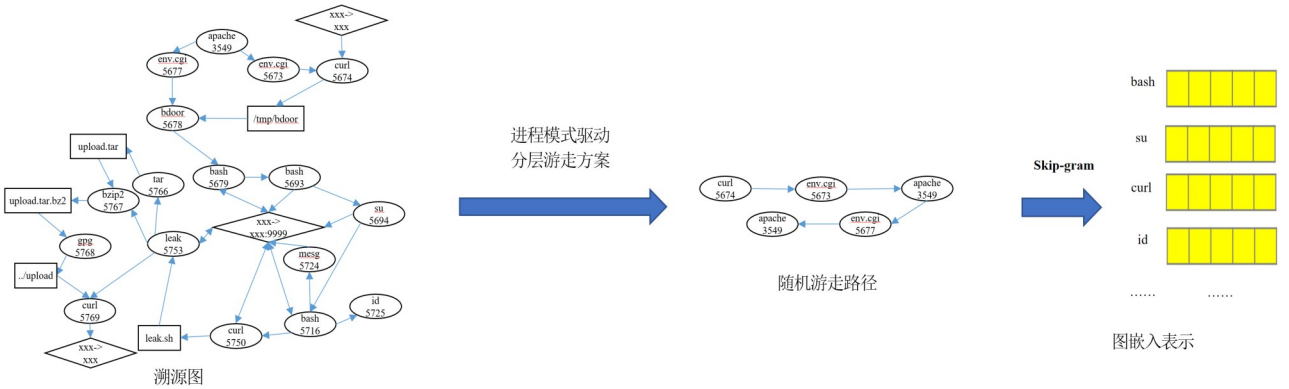

图13 基于分层游走的溯源图序列表征方法流程图

4.1.2 基于子图的表征方法

子图[74,75]能够表示攻击行为，基于序列的表征方法表达了攻击或进程模式，但序列只能从一阶邻居节点处获取信息，在嵌入的时候会丢失高阶的语义信息，因此难以获取攻击行为语义表达．用户行为在溯源图中可以抽象为数据流和控制流，可以通过跟踪事件之间的信息流[76]来识别用户的行为．行为是指用户为实现目标而执行的一系列操作，因此从溯源图中识别行为可以转化成子图提取[77,78]问题．在获取行为对应的子图之后，通过图嵌入[79]的方式来获取对应的语义信息．针对提取出的行为进行分析，能够更加深入地理解正常行为与攻击者行为的差异．WATSON[80]从数据对

象(即文件和网络连接)开始,到可能发生依赖爆炸的节点结束,采用改进的深度优先搜索算法[81]获取用户行为子图,进行图嵌入获得实体和关系的表示. 与行为无关的事件在子图中更为普遍,它们在不同的行为中重复出现,而与实际行为相关的事件发生的频率较低,根据逆文档频率算法[82]将实体和关系的嵌入聚合成行为的嵌入表示,基于子图表征方法表达了行为. 在构建行为嵌入表示的过程中,如何合理地为实体和关系赋予权重,以充分表达其对行为模式的重要性很关键. 正常行为与攻击行为的表现形式多样且数据分布极度不平衡,对于低频发生的攻击行为,如何有效提取其子图并形成具有代表性的嵌入表示,是重要的研究痛点.

4.2　基于多维语义表示学习的攻击调查方法

根据学习到的攻击模式、进程模式和用户行为等表达,可以辅助或完成攻击调查. ATLAS[59]将攻击调查转化为分类问题[83],针对每个待检测实体,将其和所有已知攻击实体构建为一个集合并提取序列,将序列进行嵌入,输入到包含攻击模式信息的LSTM模型中判断待检测实体是否为攻击实体. 还原出所有的攻击实体之后,通过抽取这些攻击实体的一阶邻居来重建攻击场景,完成攻击调查. 但ATLAS需要对每个待检测实体进行序列嵌入和模型判断,计算成本可能较高. DEPCOMM[68]为每个进程社区生成一个摘要图,概括出社区中最可能与攻击相关的进程活动,从警报事件、事件类型、事件频率和时间跨度等多个维度,为社区中的路径评分,选取评分最高的几条路径组成社区摘要图. 通过分析社区摘要展示的进程行为,安全分析人员能够迅速定位还原攻击场景,实验结果表明每个社区中排名前二的摘要路径就足够覆盖整个攻击场景,大大减少了安全人员工作量,加快应急响应处置速度. 但社区路径评分需要考虑多个维度,可能需要合理权衡不同因素,对评分的准确性依赖程度高. WATSON[80]对提取的子图行为实例进行层次聚类分析[84](Hierarchical Clustering Analysis, HCA),最初,每个行为实例都属于单独一个类,HCA算法计算类之间的余弦相似度[85],并组合两个相似度最高的类,直到所有类间相似度都小于给定阈值,其中相似度基于两个类算术平均值进行计算. 确定了聚类的簇之后,通过计算每个实例与其余实例的平均相似度来量化簇中每个实例的代表性,并根据簇的代表性实例提取行为签名. 图14展示了层次聚类算法的具体流程. 对获取到的用户行为实例聚类之后,没有良性行为被错误地聚类到攻击行为之中,这说明了行为抽取在减少分析工作量的同时,能够保证攻击调查的准确性,从溯源图中抽取行为这一方法可以辅助攻击调查. 但在攻击场景中,可能存在误导性信息或噪音,这可能会影响方法的有效性和准确性,如何处理这些误导性信息是一个挑战. 未来的研究方向可以聚焦于改进模型以应对未知威胁、优化特征表示来增强对攻击行为的理解,开发更加稳健且具有自适应性的聚类和推理算法.

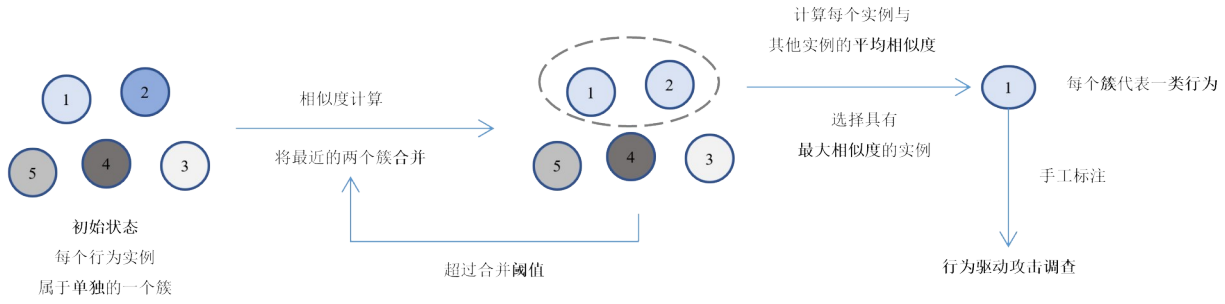

图14　基于层次聚类分析的溯源图攻击调查方法流程图

4.3　小结

基于深度表示学习的攻击调查方法通过挖掘溯源图中的隐式特征辅助攻击调查,以序列或子图的形式对溯源图攻击行为进行表征,训练模型学习攻击模式,进而发现攻击线索,还原攻击场景. 基于深度表示学习的方法面临一些关键问题,在攻击调查中,特征的可解释性对于理解攻击行为和分析威胁非常重要,因此,关键问题之一是如何提高深度表示学习模型的可解释性,使得安全专业人员能够理解模型的预测和决策. 在实际环境中,攻击调查通常需要实时响应,但深度表示学习模型的计算复杂度可能较高,如何在保持高效性能的同时满足实时性的要求是亟待研究的课题. 训练深度表示模型需要大量的标注数据,网安领域数据存在显著的数据类别不平衡问题,模型能力受到数据限制,难以发现新型攻击行为,因此深度表示学习方法捕获攻击线索能力不足. 研究者们通过引入异常检测方法,利用深度表示方法挖掘的隐式特征,在不依赖标注数据的前提下学习正常行为模式,发现异常攻击行为,解决上述问题.

5　基于异常检测的攻击调查方法

异常检测[86,87]是攻击检测的常见方法,通过综合

利用显式和隐式语义特征，发现异常攻击行为，提供更多线索供攻击调查使用．异常检测的思想是对系统正常行为建模，偏离正常状态的行为即为异常．所以异常检测方法对标注数据的需求较低，有效地缓解了网安领域标注数据缺失的问题．在不需要专家知识的情况下，检测未知攻击．早期基于溯源图的异常检测研究主要通过统计方法实现，基于正常事件建立良性数据库，通过事件频率等显式特征量化异常度，该类方法没有利用溯源图丰富的上下文语义，只关注单个事件的统计特征，检测效果不佳．近年来，相关工作大多和先进的深度表示方式相结合，在显式特征基础上，通过向量化溯源图，结合聚类和图神经网络等先进智能算法学习深度隐式特征，发现更多隐蔽攻击线索，迅速关联完成攻击调查．异常检测的基础是确定待检测对象，通过提取溯源图中的语义单元，表达成子图或路径的形式，对其进行异常度评估，发现攻击行为．基于异常检测的攻击调查方法可分为基于子图分析和基于路径分析的方法．图15展示了基于异常检测的攻击调查方法分类以及相关重要文献．

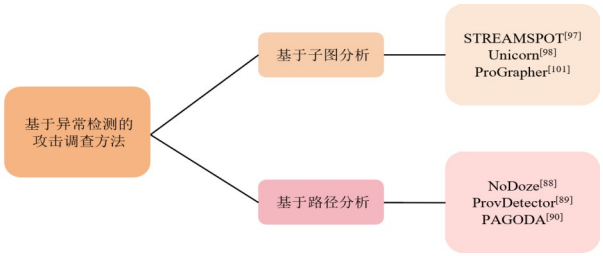

图15 基于异常检测的攻击调查方法分类图

基于异常检测的攻击调查方法首先会将溯源图拆解成待检测对象，一般表达为路径或子图，使用异常分数或者向量来描述其特征．最后通过统计或学习方法将拆解的语义单元转化为可度量形式，基于度量结果及溯源图语义特征完成异常语义单元的检测．图16展示了基于异常检测的攻击调查方法的一般性框架．

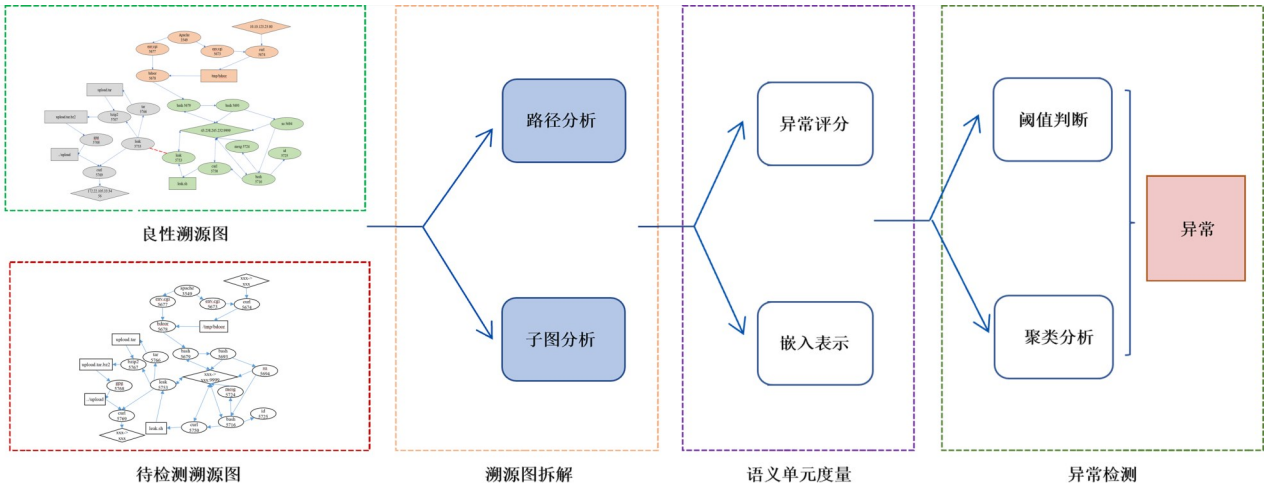

图16 基于异常检测的攻击调查方法框架图

5.1 溯源图拆解方法

5.1.1 基于路径分析的拆解方法

在异常检测系统中选择路径作为检测对象，一方面是由于用户行为在溯源图中表现为操作序列，另一方面，隐蔽的现代网络攻击企图留下最少的痕迹从而逃避检测，导致审计日志记录中大部分活动都是良性事件，只有将溯源图拆分成更小单元，比如路径，才能够更好地将溯源图的良性与恶意部分区分开．NODOZE[88]以警报点为基准，获取包含警报点的固定长度依赖路径，每条路径都包含警报的祖先和后代因果事件，表示导致警报点的事件链和由警报点引发的事件链．根据警报点生成依赖路径，能够将调查范围聚焦到潜在攻击事件中，避免对全图进行分析，产生路径依赖爆炸问题．但该方法依赖于警报点的准确性，可能受到警报点选择的影响．PROVDETECTOR[89]认为被劫持控制的隐蔽恶意程序往往需要依托良性进程或正在运行的进程来执行后续恶意活动．判断目标程序是否被劫持的问题，可以转化成监测目标程序创建的进程实例活动是否存在异常的问题，对于每个程序，从进程启动到进程结束，记录该过程中进程的所有操作和交互行为，称之为一次进程实例活动，从而生成代表进程行为的依赖路径．但如果恶意程序能够模仿正常进程的行为，或者劫持了合法进程来执行恶意活动，那么这种方法就容易受到欺骗．且恶意活动可能涉及多个层次和阶段，可能需要综合考虑不同层次的信息才能进行准确的检测．单纯监测进程实例活动可能无法提供足够的上下文综合分析．PAGODA[90]从溯源图中所有入度为0的节点出发，搜索所有路径．Log2vec[91]使用随

机游走算法在异构图中生成依赖路径,日志中包含主体、客体、操作类型、时间和主机等信息维度,只考虑日志本身会忽略日志时序约束,难以获取用户变化的行为习惯. 根据设计的启发式规则,在日志时间序列数据构建的异构图上进行随机游走,保存反映用户典型行为并暴露恶意事件的信息. 启发式规则分为三个维度:一日内的日志序列时序关系、多日内的日志序列间逻辑关系以及主机间交互关系. 由于方法中未考虑用户行为变化的监测,在长期使用中可能无法随用户行为演化而演进. 图神经网络[92,93]作为近年来异构图表示学习的重要方法,许多研究者将其引入研究溯源图上的异常检测问题,图神经网络的本质是消息传递,将邻域信息聚合到节点上,对应到溯源图上,节点的行为(操作或被操作)信息被聚合,可以理解为将溯源图拆解为以节点为中心的依赖路径. IPG[94]提出了一种基于元路径聚合的图神经网络方法对溯源图节点进行表示. SHADEWATCHER[95]认为仅提取低层的邻居信息不足以表达节点的依赖路径语义,所以对溯源图进行增强,通过提取高阶连接的方式来补充实体语义信息,由于溯源图规模一般都很庞大,因此直接进行长距离依赖提取在实际操作中是不可行的. 根据图中的资源节点,延伸出符合时序关系的依赖路径,路径上的节点可以添加长距离依赖,从而丰富了图神经网络信息聚合的语义层次. 但长距离依赖可能导致模型更加复杂,增加训练和推理的计算开销,因此需要权衡模型的复杂性和性能. 检测过程中,路径的选择和定义可能涉及到多个方面,包括长度、元素的组合方式等,不同的选择可能会导致不同的异常检测性能.

5.1.2　基于子图分析的拆解方法

语义相关的行为在溯源图中往往不是严格意义上的序列,而是包括许多分支和交互行为的子图,依赖路径序列语义单元虽然可以反映行为的依赖关系,但缺乏攻击事件的完整上下文信息,且依赖路径的长度不好控制,易遗漏关键信息. 因此,利用图可以更好地学习攻击行为特征来辅助异常检测. STREAMSPOT[96]将图作为检测对象,UNICORN[97]将一定时间内日志的增长量作为溯源图子图划分的依据,将固定时间间隔收集到的完整溯源图不断转化为直方图,使用WL子树图核算法[98]获得溯源图中的拓扑信息,分析图的结构状态变化,借助这种结构信息来判断异常状态. 如图17所示,WL子树图核算法聚合溯源图中每个节点的周围数跳邻居信息,使用哈希算法生成新的节点标签,然后统计每种节点标签的数量并生成直方图,最终生成的直方图表示了溯源图的结构信息,压缩了攻击行为信息,检测的对象是新增长的部分图. PROGRAPHER[99]将溯源图构建成一系列时序快照,每个快照包含固定数量节点,相邻快照之间存在部分重复节点,以快照作为异常检测的对象. 上述方法都将溯源图信息进行了一定的压缩,但在信息压缩的同时,如何确保不会丢失重要的攻击行为是需要仔细考虑的问题.

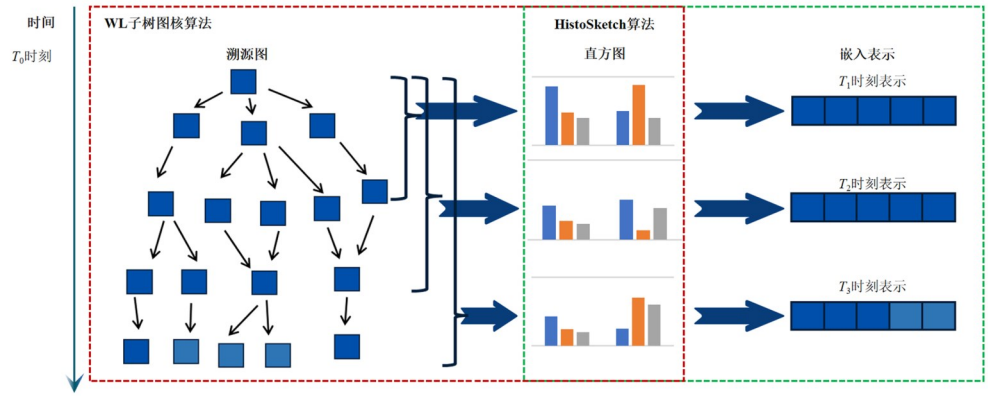

图17　基于WL子树图核算法的图拆解方法流程图

5.2　溯源图语义单元度量方法

5.2.1　基于异常评分的语义单元度量方法

要实现对溯源图语义单元数值化有一个朴素的思路,即对语义单元进行异常评分. NODOZE[88]通过异常分数来量化溯源图中依赖路径的可疑程度,首先统计路径中所有事件在良性数据中的发生频率,事件发生的频率越低,越可能是异常事件;路径的异常分数计算不仅考虑到每个事件的发生频率,还考虑了事件所处源目节点的图结构特征. 将节点运行的时间范围切分成N个时间窗口,将入边分数定义为有入边到达源节点的时间窗口比例,将出边分数定义为目的节点有出边的时间窗口比例,他们分别表示实体作为信息接收者和发送者的重要性. 将发生频率与入边和出边分数相乘,得到事件异常分数,最终将路径中多个事件的异常分数连乘,计算出路径的异常分数. 入边和出边分数的定义涉及到节点作为信息接收者和发送者的重要性,这种定义可能受到具体场景的影响. 不同的场景可

能需要不同的度量方式，而通用的定义可能无法适应所有情况. PAGODA[90]维护了一个正常行为数据库，为了准确跟踪和获取进程正常行为的溯源信息，PAGODA 记录了多次进程正常行为，并设置一个阈值 T，统计审计事件在进程运行中出现的次数 N. 如果 $N>T$，我们认为对应的统计事件是正常事件，并将其放入规则库，通过统计路径中不属于正常事件规则库的事件比例，得到依赖路径的异常分数. 基于评分的方法通过统计路径中不属于正常事件的事件比例来计算异常分数，这种计算方式可能无法充分考虑事件的重要性和上下文信息，导致对异常的判断不够精准.

5.2.2 基于嵌入表示的语义单元度量方法

利用异常分数对语义单元进行度量难以充分挖掘单元的语义，缺乏对溯源图事件序列或子图的上下文理解，所以研究者们尝试引入先进的机器学习智能算法对语义单元进行表示. Log2vec[91]利用随机游走和 word2vec 算法[100]将异构图嵌入为向量，原始随机游走算法在采样过程中会导致恶意行为与良性行为关联，降低日志条目之间的差异. 使用改进后的随机游走算法，根据事件的统计规律将主机的用户行为划分为多元场景，针对不同的场景设定不同的权重策略，从中选择权重较大的边，进行随机游走，产生与之对应的行为序列，作为语料库进行词嵌入，将细粒度的日志项转化为数值型的高维向量. 生成的语料库（行为序列）的质量对 word2vec 的嵌入效果有很大影响. 由于该方法基于随机游走，语料库的质量可能不稳定，从而影响最后的向量表示. 且随机游走过程可能会引入噪声，混淆正常行为和异常行为的边界. 直接对图进行拆解分析，会产生大量待检测依赖路径，PROVDETECTOR[89]使用基于稀有度的路径选择算法来识别溯源图中的部分可疑路径，这些路径代表了进程的潜在恶意行为，降低计算代价. 借鉴先前工作 NODOZE[88]中的异常分数计算方法，PROVDETECTOR 对每条依赖路径打分，只选择前 20 条异常分数最高（最可能是潜在的异常路径）的路径进行后续的训练和检测. 要将这些路径提供给异常检测模型，需要将它们嵌入到高维向量空间，向量在高维空间中的位置代表它们的语义信息. 通过将依赖路径视为句子，利用 PV-DM 模型[101]将每条依赖路径嵌入成高维向量. 其中，PV-DM 模型联合训练段落向量 PV 和单词向量 DM，使得模型能够更好地捕捉文本的语义信息. 同时它可以学习任意长度词序列的分布式表示，并且考虑到了单词之间的顺序，这对有序攻击步骤分析至关重要. PV-DM 模型在处理安全领域特定术语、API 调用序列等时，其预训练模型可能需要针对此类数据进行微调以提高准确性. THREATRACE[102]提取溯源图中节点所有的出入边相关信息作为节点特征，使用 GraphSAGE 框架[103]进行聚合. SHADEWATCHER[95]在对溯源图的语义信息进行增强之后，采用 TransR[104,105]对系统实体和边的表示进行学习，在学习的时候通过对正常的实体交互进行变换，随机替换正常交互中的实体生成异常实体交互，作为反例来进行对比学习. 在获取实体和边的向量表示之后，采用图神经网络对实体的上下文语义进行补充. 不同的实体对当前实体的贡献不一样，采用注意力机制[106,107]学习邻居实体的重要性，利用重要性对相邻实体加权，对相邻实体的信息进行聚合，优化实体的表示. 但随机替换正常交互中实体生成异常实例作为反例进行对比学习的方法依赖于所构建异常数据的质量，如果替换过程无法有效模拟真实世界中的异常模式，可能会影响模型的学习效果. UNICORN[97]将溯源图转化为直方图，但是直方图语义离散，是描述系统执行情况的简单向量空间图形统计，无法匹配攻击检测需求. UNICORN 采用 HistoSketch 算法[108]将直方图转化为固定大小的向量，压缩溯源图中包含的海量信息，将细粒度的溯源图抽象到高维度的向量表示. 但固定的向量大小可能导致更高程度的信息损失，特别是对于异常行为或恶意活动的微妙特征可能无法完全保留，系统更容易产生误报和漏报. PROGRAPHER[99]生成的快照是复杂的图结构表示，利用 graph2vec 算法[109]从快照中提取有根子图并基于 skip-gram 算法对快照的向量做表示学习. 基于嵌入表示的语义单元度量方法能够捕获攻击事件序列的时序关系，嵌入的向量将节点和边的信息表达到一个维度更小的向量中，基于向量的计算方法也更加丰富，更好地利用溯源图的特征. 然而对于大规模的图或实时应用，这些方法的计算效率可能是一个问题. 此外，由于这些方法将图中的节点和边转化为高维向量，降低了模型的可解释性，使安全分析师难以直接从向量中还原出原始的系统执行路径和实体间的关系.

5.3 异常检测方法

5.3.1 基于聚类分析的异常检测方法

获取了待检测对象的嵌入向量表示后，可以采用聚类分析的方法来发现偏离正常的行为. PROVDETECTOR[89]使用局部异常因子算法对依赖路径的特征向量进行检测，局部异常因子算法[110]假设非离群点对象周围密度与其邻域周围密度类似，而离群点对象周围密度显著不同于其邻域周围密度，特征向量在数据集中密度越低，越有可能是异常路径. 但如果真实数据中存在大量的噪声或数据分布极度不均匀，那么基于密度的方法可能会受到干扰，导致误判或遗

漏真正的异常路径. 如图 18 所示,UNICORN[97]通过对系统良性工作模式进行建模来检测异常状态. 利用良性溯源图构建进化模型,基于训练期间创建的一系列按时间顺序排列的向量,使用 K-medoids 算法[111]对特征向量序列进行聚类,由此建立进化模型. 簇代表工作阶段模式,并且簇之间存在时序关系,使用簇中特征向量的时间顺序以及簇的统计数据来生成系统演化的模型. 进化模型包含了系统运行时执行状态的变化. 通过构建进化模型,描绘企业中工作模式,获得高置信度良性工作场景,待检测向量如果不在聚类形成的簇中或是不符合簇之间的进化趋势,则被认为是异常行为. 该方法考虑了簇内聚合、簇间演化的特点,具有一定的自适应性,然而描述系统工作状态之间时序关系的复杂度较高,尤其是当系统行为具有非线性、非平稳特点时,简单的聚类分析难以完全捕捉到这种复杂的时间依赖结构.

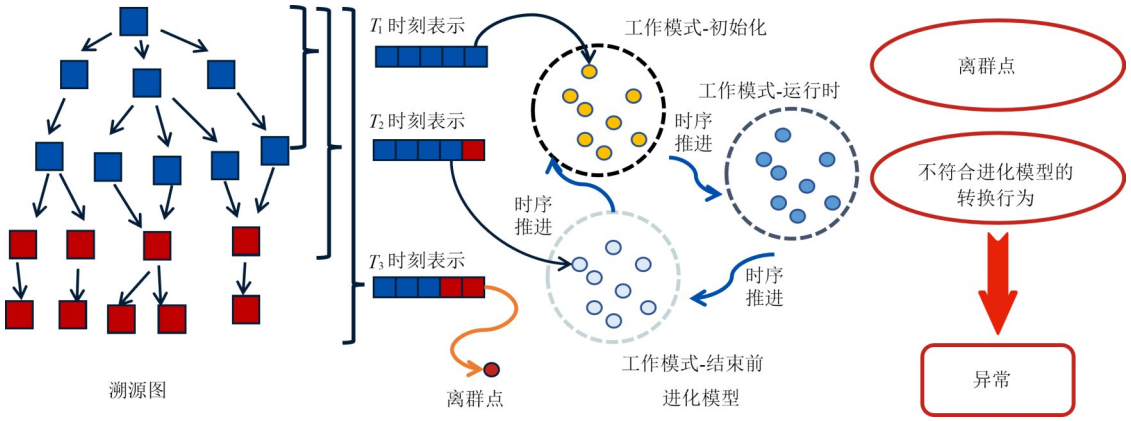

图 18　基于进化模型的异常检测方法示意图

5.3.2　基于阈值判断的异常检测方法

基于聚类分析的方法适用于大规模向量的分析,许多工作将异常检测问题进行转化,分析结果是异常程度的数值表示,通过阈值判断的方法可以较好地对数值结果进行分析. NODOZE[88]和 Pagoda[90]认为良性路径与异常路径的异常分数之间存在数量级的差异,在进行异常检测时,引入阈值来量化两者之间的差异,通过在包含真实攻击的数据集上调整阈值,平衡准确率和虚警率. 这种方法能够较为直观地衡量系统状态偏离正常行为的程度,根据真实攻击场景调整阈值,达到良好的检测效果. Log2vec[91]使用无中心聚类算法对日志嵌入的行为特征向量进行聚类,如果聚类后的簇中,日志数量小于阈值,该簇被视为异常簇,根据异常簇即可发现高可疑攻击者. 然而在出现大量恶意日志的情况下,簇内日志数量可能较高,导致检测失效. SHADEWATCHER[95]发现了推荐和异常检测任务的相似性,将推荐系统思想引入到异常检测领域. 通过对异常检测和推荐系统[112]进行比较,将系统实体之间的交互映射到推荐系统中用户和项目之间的交互;同时根据推荐领域中行为相似的用户具有相同的偏好这一假设,提出了语义相似的实体具有相似的交互偏好,将异常检测问题转换为预测实体之间是否交互的问题,并参考推荐中的辅助信息,从上下文中获取实体的语义信息来对应推荐系统中的辅助信息进行异常检测. SHADEWATCHER 的嵌入方法保证正常实体之间的交互在实体进行点乘的时候具有较小的值. 如果两个实体点乘结果超过了给定的阈值,则认为这两个实体之间不应该存在交互,但是实际上它们已经产生了交互,此时就认为这两个实体之间产生的交互是异常行为. 与推荐系统类似,SHADEWATCHER 考虑了上下文信息对实体交互的影响,增强了模型对于环境变化、时间序列等因素的理解能力,提高了异常检测的准确性. 但并非所有实体关系都能简单映射到向量空间中并保持原有语义结构不变,如何保证实体嵌入在点乘运算后有效地区分正常和异常交互是一个问题. PROGRAPHER[99]通过对快照序列生成的向量进行训练,利用 TextRCNN[113]进行预测,当新的待检测快照到来时,将预测的嵌入向量与真实嵌入向量进行比较,如果它们之间的距离超过预定义阈值,将其标记为异常. 基于阈值的异常检测方法可以清晰定义正常和异常之间的边界,通过调整阈值来适应不同场景下的准确率和虚警率要求,但如何设置一个既能有效区分异常又不过于敏感的阈值可能颇具挑战.

5.4　小结

基于异常检测的攻击调查方法能够融合基于因果分析和基于深度表示学习两种方法的优势,结合它们表征溯源图的方法,将溯源图转化为可度量的对象,并通过学习良性行为来判断系统异常状态,为攻击调查提供更多的线索. 异常检测可以检测未知攻击,并且不需要专家知识. 但同时,基于异常检测的方法会导致与

正常行为不一致的事件全部被判定为异常,然而在现实场景中异常行为不都是攻击.在实际应用中,正常行为数据可能是不确定的或不完整的,如何在缺乏明确标签的情况下有效地进行异常检测是一个关键问题.异常检测模型学习的系统正常行为是时间敏感的,随着时间推移,可能发生工作模式变化等概念漂移现象,导致误报率较高.所以如何加深对攻击的理解,提取更有效的行为表达特征,在保持高敏感性的同时,有效地控制误报率,以提高调查方法的可用性极为关键.表1总结了基于异常检测的攻击调查领域的代表性研究成果,从模型/算法、调查流程、数据集及研究特点等几方面做了对比研究.

表1 基于异常检测的攻击调查方法小结

方法	模型/算法	图拆解方法	语义单元度量	异常检测	数据集	特点
NODOZE[88]	深度优先搜索算法	基于警报点的定长依赖路径获取	异常分数:罕见度特征、图结构特征	异常分数阈值判断	NEC Labs America191台主机上5天内收集数据(包含10次模拟APT攻击)	评估警报异常分数降低误报率
PROVDETECTOR[89]	PV-DM模型+局部异常因子算法	进程启动到结束的所有操作和交互行为	训练段落向量PV、单词向量DM捕获有序攻击步骤语义信息	特征向量密度越低异常概率越高	某企业内网150台主机上一周内收集数据(包含8个攻击案例)	针对恶意软件行为进行检测
PAGODA[90]	深度优先搜索算法	从图中所有入度为0的节点出发搜索所有路径	维护良性行为数据库统计异常事件比例	异常分数阈值判断	自主收集的8个攻击场景数据	建立良性行为数据库优化图存储格式
Log2vec[91]	随机游走+word2vec+聚类	根据启发式规则随机游走获取单日内、多日间关系	word2vec算法将游走的路径转化为向量	无中心聚类根据簇内向量数量进行阈值判断	CERT数据集、LANL数据集	检测企业内部用户异常行为,分析反映用户典型行为的信息,发现恶意事件
THREATRACE[102]	GraphSAGE+多模型框架	以节点为检测对象	GraphSAGE聚合节点两跳邻居信息形成特征向量	多模型框架将无法正确分类的节点视作异常	StreamSpot数据集、UNICORN数据集、DARPA数据集	学习节点行为模式,用有监督的分类任务替代无监督的异常检测任务
SHADEWATCHER[95]	TransR+GNN	为节点添加长距离邻居信息,形成依赖路径	TransR对实体和关系进行嵌入	GNN计算实体间是否应该存在交互关系	DARPA TRACE数据集、自建模拟数据集(包含6个攻击场景)	利用推荐系统思想完成高效异常检测
STREAMSPOT[96]	StreamHash+K-medoids	以图为检测对象将溯源图分割成小块	StreamHash算法将图映射为向量	利用K-medoids算法对向量进行聚类,距离簇较远的为异常图	StreamSpot数据集	图级别的高效流式异常检测
UNICORN[97]	WL子树图核算法+HistoSketch算法+聚类	一定时间内的增长图作为检测对象,并转化为直方图	HistoSketch算法将直方图转化为向量	K-medoids算法聚类并形成进化模型,表达工作模式	UNICORN数据集、DARPA数据集	图级别检测,以图增长部分作为检测对象,提高精度并缩小调查范围
PROGRAPHER[99]	graph2vec+TextRCNN	将流式图的快照序列作为检测对象	graph2vec算法将快照序列嵌入成向量	训练生成模型TextRCNN,计算模型嵌入和真实嵌入的差距,检测异常	StreamSpot数据集、DARPA数据集、ATLAS数据集、企业EDR数据集	结合整图嵌入和序列学习,分析溯源图快照,学习了正常系统的行为表示

6　溯源图优化方法

在基于溯源图的攻击调查系统中,溯源图的存储计算是基础. 研究者们发现,许多先进方法应用到真实企业内网环境中,性能并不理想. 由于现代软件系统十分复杂,且网络攻击时间跨度大,导致溯源图规模增长快,调查攻击全貌所需的日志覆盖范围广. 在推进基于溯源图的攻击调查系统应用落地过程中,最大的瓶颈就是溯源图庞大规模带来的存储和计算压力. 为了突破这一难题,研究人员提出溯源图优化方法,从多个维度缓解存储计算压力. 早期工作集中于基于有损压缩的溯源图优化方法,在尽量不影响攻击系统行为的前提下,删除冗余节点和边,对图进行剪枝. 这类方法能够大大缩减溯源图规模,减轻存储压力,更重要的是能够为攻击调查排除大量干扰,加速因果分析,提高调查效率. 然而在剪枝过程中难以避免攻击相关信息的丢失,随着计算能力的发展,研究者们开始探索无损压缩方法,保留全量攻击信息,针对溯源图结构特点、溯源图属性对象特征,提出对应的合并和编码方法. 最近的方法引入深度神经网络,将溯源图向量化,形成溯源图向量数据库,极大压缩存储空间,提高查询效率. 溯源图优化方法可以分为基于有损压缩和基于无损压缩的方法. 溯源图优化方法的分类和典型文献如图 19 所示.

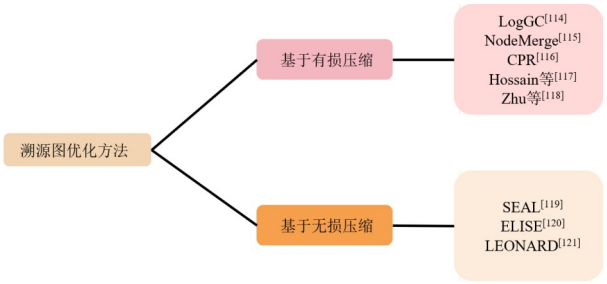

图 19　溯源图优化方法分类图

6.1　基于有损压缩的溯源图优化方法

基于有损压缩的方法通过对溯源图进行结构优化,去除冗余节点、边和图上环路,提高图计算性能,在保留尽可能完整因果语义的前提下,减小溯源图规模. Lee 等[114]提出的 LogGC 方法类似于堆内存管理,旨在通过分析收集的日志数据来移除被认为对因果关系分析无影响的临时文件节点. 然而,此举可能导致删除与攻击相关的临时文件,从而丧失可能影响攻击调查工作的攻击信息. NodeMerge[115]通过自动学习固定库和程序的只读资源集作为模板,在进程初始化期间合并只读事件,减少冗余节点,然而,攻击者只需让恶意软件在实际攻击之前长时间潜伏,即可打破因果依赖关系,在 APT 攻击场景下,这种情况是完全可能的. APTSHIELD[24]删除执行了退出事件且超过五分钟没有活动的节点. 但正常情况下系统日志生成的溯源图是一个稠密图,其边的数量远远大于节点的数量,所以针对节点的优化剪枝算法效果比较一般.

要想达到良好的优化效果,需要针对溯源图的事件边进行剪枝. CPR[116]聚合依赖性相同的事件,保留图的网络拓扑结构,但会丢失时序上的因果信息以及访问频率等统计信息. 要在保证因果可达性的前提下剪枝,需要计算图上节点之间可达性等全局图属性,由于溯源图是带有时间属性的图,随着时间发生变化,缓存计算的全局图属性会失效,大大增加了图计算成本. 如图 20 所示,Hossain 等[117]基于版本化思想,将溯源图转化为一个标准图,将时间戳从边转移到节点上,将时序上的可达性转换为节点之间的可达性,使得溯源图支持缓存计算. 每次节点获得新的入边时,都会创建该节点的新版本,针对节点或边,进行因果影响分析,将不带来新语义的节点或边进行融合处理,使得溯源图规模可控,降低图上计算成本,但是这类方法是针对因果分析的特化处理,适用范围较窄.

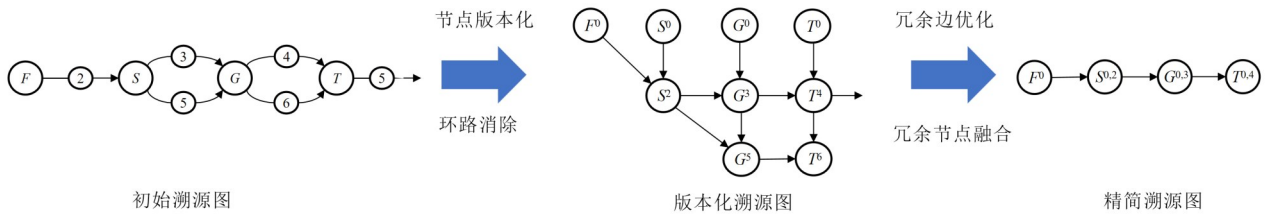

图 20　基于版本化剪枝融合的溯源图增强方法流程图

DEPIMPACT[45]提出了局部溯源图生成方法,从警报点开始,执行反向因果分析,搜寻与警报点有直接因果路径的节点和边,删除不满足时间约束的冗余边,从而得到与警报事件高度相关的溯源图. 此溯源图旨在调查入口节点,在尽可能保留所有攻击路径的前提下,对溯源图进行修剪压缩. 且操作系统通过将数据按比例分配到多个系统调用来完成文件的读/写任务,所以 DEPIMPACT 合并时间窗口内节点间的重复边. 但是该方法生成的溯源图无法对警报点之后的攻击步骤进行分析还原,难以发现攻击的完整影响. 上述方法虽然从攻击调查任务出发,但缺乏对威胁语义的理解应用,Zhu 等[118]提出基于可疑语义的剪枝方法,将攻击者的

攻击行为抽象成可疑语义的标签传播,保留造成可疑语义传播的事件,将攻击无关事件丢弃. 与基于标签的因果分析调查方法类似,该方法依赖规则的完备性,未被可疑语义发现的事件不一定与攻击无关,导致攻击场景缺失.

6.2 基于无损压缩的溯源图优化方法

无损压缩的基本思想是使用更短的编码来代表频繁出现的元素. 无损压缩方法的主要目标是缓解存储压力,复杂的计算机系统每天会产生大量日志,例如开启所有日志记录的火狐浏览器,每天将产生10 GB以上的日志,这对服务器负载接近上限的企业来说难以接受. 通过应用无损压缩算法,可以使存储成本变得可接受,并且能够保留溯源图的全量信息以供进一步调查.

SEAL[119]是最早针对溯源图做无损压缩的工作,从图结构和属性值两个维度发力,对溯源图进行优化. SEAL将每个节点的父节点进行合并,创建对应的映射表,确保解压缩可靠性,再对各项属性中空间占用最大的时间戳进行增量编码和哥伦布编码,在保留时序的前提下大幅减少所需存储空间. 但是SEAL压缩算法可能会导致一些事件先后顺序的丧失,可能会对一些时间限制的查询产生影响. ELISE[120]在提取溯源图属性值的冗余模式之后,引入深度神经网络模型,以模型预测概率代替算术编码表,对所有属性值进行算术编码,解决了编码需要先验知识的问题. LEONARD[121]在此基础上更进一步,将整个溯源图向量化,构建图向量数据库. 首先将溯源图解耦成边表与节点表,将图结构展平为可编码的文本信息,再将文本使用char2vec方法转换为向量,逐字符输入到轻量级LSTM模型,训练方法为根据当前字符预测下一字符. 由于模型属于轻量级设计,面对庞大的溯源图数据时可能出现预测错误. 然而,为了确保编码的无损性,LEONARD在每次迭代结束后引入修正表,对当前模型的错误进行修正,最终以向量的形式存储数据,极大压缩了存储空间,在查询和解压时,模型能够根据关键字联想出详细信息,查询效率同样可靠. 然而基于神经网络模型的方法运行时间相对较长,与传统的压缩方法相比,模型推断更耗时. 在对存储数据进行更新时,此类方法操作效率低下,因为每当存储的数据发生变化时,模型就需要重新训练,增加额外开销.

6.3 小结

溯源图优化方法是构建基于溯源图的攻击调查系统的关键基础. 有损压缩方法在减轻存储压力的同时,更加注重对攻击无关事件的剪枝处理,提升攻击调查效率;无损压缩方法专注于溯源图存储空间的压缩,相关方法针对溯源图的特殊结构、时空因素提出独特的压缩算法,但是压缩时带来的开销以及全量、海量数据对攻击调查效率的影响同样需要关注. 因此,如何结合攻击语义,提高攻击相关事件解压缩和调查速率,尽可能去除冗余节点和边,达到调查效率和攻击场景完整性的平衡,是溯源图优化方法需要突破的关键问题.

7 实验分析

完成了对各种攻击调查方法的介绍,本节通过分析比较基于溯源图的攻击调查方法的实验内容,包括数据集、评价指标等,进一步提升对攻击调查任务的认识. 由于所有类型方法直接进行横向对比十分困难,本文选取四种基于异常检测的攻击调查典型方法STREAMSPOT[96]、UNICORN[97]、PROGRAPHER[99]和THREATRACE[102]进行横向效果对比. 最后对三类攻击调查方法的优劣势进行详尽总结分析.

7.1 数据集介绍

基于溯源图的攻击调查领域常用的数据集主要包括DARPA透明计算数据集[122]、STREAMSPOT数据集[96]和各类自建数据集.

(1)DARPA透明计算数据集

DARPA透明计算数据集由5个子数据集组成,包括Trace、FIVEDIRECTION、CLEARSCOPE、THEIA和CADETS,它们是在DARPA透明计算计划下构建的. 每个子数据集包含特定系统事件(例如,文件读/写、网络连接)在各种操作系统平台(如Windows、Linux和FreeBSD)上的系统日志,时间跨度为两周. 红队在“沉默”期(仅执行良性活动)后执行多次攻击活动. 攻击活动模拟已知的APT攻击,如Nginx后门、Darkon APT、Firefox后门,以及常见的攻击,如发送钓鱼电子邮件等.

(2)STREAMSPOT数据集

该数据集包括基于1次攻击和5种良性场景生成的溯源图. 良性场景涉及正常的浏览活动,包括观看YouTube、下载文件、浏览网站、检查电子邮箱和玩电子游戏等. 攻击行为是通过访问恶意URL触发的驱动下载,利用Flash漏洞获取主机的管理员权限. 每个场景会运行100次,总计生成600张图.

(3)自建数据集

为了更好地理解攻击行为特征,许多论文通过模拟攻击或与企业合作来自建数据集. ATLAS数据集[59]是在实验室环境中收集的,包含10种APT攻击,包括钓鱼邮件和横向移动等不同策略. 在每次攻击执行期间,各种良性活动(包括浏览网站、阅读电子邮件、下载附件、连接到其他主机等)都一起模拟. 平均每个场景有20 088个实体和25万个事件. 与攻击相关的实体被标记为恶意. RAPSHEET[123]收集了来自赛门铁克公司内部运行的34台主机的系

统日志和威胁警报. 数据跨越一周时间,从产品开发团队成员使用的主机上收集而来. 这些主机上执行的任务包括网页浏览、软件编码和编译、质量保证测试和其他日常业务. 总共收集了约35 GB的日志,包含约4 000万个系统事件. 在实验期间,攻击者对三台主机发起了攻击,这些攻击行为对应三个不同的攻击活动,其中两个基于现实世界的APT威胁组织,另一个是自定义的数据窃取攻击. DISTDET[124]在知名安全公司生产环境中的1 130台主机(包含168台windows主机、962台Linux主机)上收集系统审计日志,在为期14天的攻防对抗中收集了约16亿个系统事件.

当前的攻击调查方法在实验设计和数据处理方面存在差异,导致难以进行直接的横向对比. 基于因果分析的调查方法常常缺乏重建效果的量化指标描述,如场景重建准确率和误报率等. 基于深度表示学习的方法多使用自建数据集,或从其他方面辅助完成攻击调查,因此上述两类方法较难横向比较分析. 因此本文实验效果分析主要聚焦于基于异常检测的调查方法. 现有异常检测工作大部分都对DARPA子数据集和STREAMSPOT数据集进行了调查分析,自建数据集一般用于进一步分析验证研究者提出方法的效果. 所以接下来的效果分析主要基于DARPA子数据集和STREAMSPOT数据集. 表2是所分析数据集的基础统计特征.

表2　实验数据集基础统计分析

数据集	平台	节点数量/个	边数量/条	数据大小/GB
DARPA THEIA	FreeBSD	**7 051 262**	358 868 880	309
DARPA CADETS	Linux	6 934 472	105 666 929	89
DARPA CLEARSCOPE	Android	699 506	**398 493 440**	**873**
DARPA FIVEDIRECTION	Windows	—	—	167
STREAMSPOT	Linux	5 046 600	89 770 900	2.8(处理后)

注:加粗数据表示该列数据中的最大值.

7.2　攻击调查评价指标

攻击调查方法的评价指标主要包括报警准确率、报警精准率、召回率、F_1值等,下面分别对它们进行介绍.

(1)报警准确率(Accuracy)

衡量了被正确报警的攻击对象在所有待检测对象中的占比.

$$\text{Accuracy}=\frac{\text{正确报警样本数}}{\text{总样本数}}$$

(2)报警精准率(Precision)

衡量所有预测为攻击对象的样本中,真正的攻击对象的比例.

$$\text{Precision}=\frac{\text{预测正确的攻击对象数}}{\text{所有预测为攻击对象的样本数}}$$

(3)召回率(Recall)

衡量所有样本中的攻击对象被正确预测出来的比例.

$$\text{Recall}=\frac{\text{预测正确的攻击对象数}}{\text{样本中所有攻击对象数}}$$

(4)F_1值(F_1-score)

精准率和召回率的调和平均数,用来衡量当样本类别不均衡时的方法性能.

$$F_1=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$$

7.3　攻击调查效果对比分析

基于因果分析的调查方法通常详细描述了其攻击场景的重建过程,并通过绘制攻击溯源场景图来呈现,但常常缺乏对重建场景的量化评估指标描述. SLEUTH[19]详细描述了还原的8个攻击场景,共计174个攻击节点,只遗漏了2个攻击节点,根据这些数据难以直接计算精准率、召回率等指标,但是能够发现SLEUTH的重建效果十分优秀. HOLMES[42]同样描述了重建的攻击场景,但只在设置虚警阈值时给出了精确率、召回率和F_1值的折线图,并未详细说明在哪些数据集上进行了实验和阈值调整. DEPIMPACT[45]调查了DARPA数据集中的5个攻击场景和自行收集的10个模拟攻击场景. 调查出的每个场景平均存在225个假阳性节点,不存在假阴性节点,同样难以计算精确率等指标. 基于深度表示学习的方法以辅助攻击调查为主,使用的数据集也各不相同. ATLAS[59]自主构建了包含10个攻击场景的数据集,平均调查精确率91.06%,平均召回率可达97.29%,平均F_1分数93.76%,性能优异. 然而ATLAS未在公开数据集(如DARPA数据集)上进行实验,无法直接对比实验效果. DEPCOMM[68]与HOLMES进行联动实验分析,在14个攻击场景中补全了HOLMES调查结果未覆盖的两个攻击步骤,但其本身作为攻击调查的辅助手段,无法直接与其他调查方法进行横向对比. 综上分析可知,基于因果分析的攻击调查方法和基于深度表示学习的攻击调查方法在各自场景下均表现出良好的性能,但难以进行对比分析. 因此本文主要对基于异常检测的攻击调查方法实验效果进行横向对比. 表3是典型方法在数据集上的指标表现.

由于不同方法在调查的粒度上存在差异,我们对文献中针对目标粒度的性能指标进行了分析.其中,STREAMSPOT、UNICORN 和 PROGRAPHER 属于异常图调查方法,其粒度较为粗略.在实际场景分析中,仍需花费一定时间对调查得到的异常图进行进一步分析,以确定攻击行为在溯源图中的具体位置.相反,THREATRACE是一种节点级调查方法,能够精准地定位到攻击节点.通过对表3的观察,能够发现PROGRAPHER在大多数数据集上取得了显著的效果.其采用了一种先进的方法,通过对图进行预处理生成快照序列,结合整图嵌入的graph2vec方法和序列学习中的TextRCNN模型,对溯源图的结构特征进行了深入分析.这种方法能够精准、高效学习正常系统行为表示,为攻击行为检测提供了可靠的基础.PROGRAPHER的优越性表现表明在攻击调查中结合图嵌入和序列学习的方法可以取得令人满意的结果.在未来的研究中,这种方法或许可以为提高异常检测的效率和准确性提供有益的参考.UNICORN在DARPA CLEARSCOPE数据集上表现良好,该数据集采自移动设备端Android系统,可能是因为UNICORN通过构建进化模型来学习正常行为,移动端的用户行为规律性更强,更符合进化模型所描述的工作模式,从而获得了更好的效果.但上述两类调查目标为图的方法具有一定局限性,在发现异常的图之后,还需要安全人员进行一定手动分析以捕获确切的攻击行为.值得关注的是,表3中所采用的指标数据来源于PROGRAPHER 和 UNICORN 原文.在PROGRAPHER中,研究人员表示UNICORN的性能未达到其论文中的水平,原因主要有两方面:(1)UNICORN使用了未公开数据集训练模型;(2)UNICORN的实验过程中,没有强制分离训练集和测试集,导致“数据窥探”问题.PROGRAPHER研究人员进行了自行实验对比,性能指标相较UNICORN均有一定程度提升,这一问题同样值得后续研究人员关注、讨论和验证.THREATRACE作为节点级别的调查系统,在各个数据集上的性能指标并不突出,但其细粒度的攻击节点定位能够大大减少安全人员分析工作量,能够显式地表达攻击相关信息,有助于分析人员对攻击进行理解,支撑未来网络防御加固.

表3 实验数据集上攻击调查方法性能指标对比

数据集	方法名称	调查粒度	Precision	Recall	Accuracy	F_1
DARPA THEIA	UNICORN[97]	图	1.00	1.00	1.00	1.00
	THREATRACE[102]	节点	0.87	0.99	—	0.93
	PROGRAPHER[99]	图	**1.00**	**1.00**	**1.00**	**1.00**
DARPA CADETS	UNICORN	图	1.00	1.00	1.00	1.00
	THREATRACE	节点	0.90	0.99	—	0.94
	PROGRAPHER	图	**1.00**	**1.00**	**1.00**	**1.00**
DARPA CLEARSCOPE	UNICORN	图	**0.98**	**1.00**	**0.99**	**0.99**
	PROGRAPHER	图	0.80	1.00	0.93	0.89
DARPA FIVEDIRECTION	THREATRACE	节点	**0.67**	**0.92**	—	**0.80**
STREAMSPOT	STREAMSPOT[96]	图	0.72	1.00	0.69	0.75
	UNICORN	图	**0.95**	0.97	**0.99**	**0.96**
	PROGRAPHER	图	0.90	**1.00**	0.94	0.94

注:加粗数据表示某数据集上各指标的最大值.

7.4 各类攻击调查方法优劣势分析

本文将基于溯源图的攻击调查方法分为三类:基于因果分析的攻击调查方法、基于深度表示学习的攻击调查方法和基于异常检测的攻击调查方法.随着先进智能算法的发展,越来越多深度学习异常检测算法被引入该领域中,但基于因果分析的方法仍是攻击调查的基础手段.因此,本文针对三类方法的优势和不足展开分析.

(1)基于因果分析的攻击调查方法

优势:该类方法通常基于一整套网络安全专家设定的规则,如标签规则、TTP规则等,能够精准发现攻击行为,并通过溯源图上因果分析将多步攻击关联起来,形成完整攻击场景.基于规则的方法具有较低的误报率,同时能够将溯源图的底层系统行为语义映射到攻击行为层面,具有很强的可解释性.这对于了解攻击者真实行为、展开针对性响应活动是至关重要的.

不足:基于因果分析的方法囿于规则限制,在遭遇新型攻击手法或面对错综复杂的攻击场景时,可能无法有效识别并精确量化威胁路径.面对规则之外的未知攻击行为容易产生漏报现象.规则的设置与更新需要高级安全专家长期维护,自适应威胁感知调查能力相对较差.

(2)基于深度表示学习的攻击调查方法

优势:现代复杂网络攻击往往包含多个步骤,步骤

之间的关联关系对于攻击调查十分重要. 基于规则的方法往往只考虑单点特征,忽略重要的上下文信息. 基于深度表示学习的方法充分挖掘溯源图中的序列、子图类型邻域信息,考虑时空、因果等关联特征,能够更有效地发现攻击行为.

不足:在基于深度表示学习的方法中,有监督学习方法依赖于标注数据,由于网安数据集中攻击数据的稀缺性,样本不平衡现象十分严重,带来的过拟合问题不容忽视. 无监督学习方法如聚类、社区发现等,其结果常常缺乏重要的攻击语义,只能用于辅助,难以直接支撑复杂攻击调查任务.

(3)基于异常检测的攻击调查方法

优势:基于异常检测的方法不依赖标注数据,有效缓解了标注数据缺乏的问题. 异常检测方法主要思想是对正常行为建模,发现偏离正常的离群行为,因此即使防御方缺乏对新型未知攻击手法和恶意软件[125,126]的理解,仍能够发现它们.

不足:在企业单位内网中,存在许多罕见但正常的操作,如更新软件、修改密码等,异常检测模型的训练数据中样本较少,可能产生误报. 同时,用户行为模式可能随时间流动而产生变化,大量新型操作可能带来许多误报,引发概念漂移问题. 如何及时更新模型,平衡攻击检测敏感度和误报率是关键的问题.

8　未来研究方向展望

8.1　未来技术路线分析

针对本文所提三类基于溯源图的攻击调查方法存在的问题挑战,本节分析并提出解决问题的可能技术路线.

(1)基于因果分析的攻击调查方法技术路线展望

针对现有方法存在的自适应威胁感知调查能力差问题,未来研究可以致力于构建适应动态场景变化的规则系统,例如根据用户行为模式,调整规则系统阈值、衰减因子等属性,使得威胁语义增强方法能够跟随环境变化而变化. 或是研究高层行为语义,发现其不变性. 攻击者的具体攻击手法可能较为新颖,其特征也会频繁变化,但攻击者的行为目标不会产生大的变化,如C2通信可能存在多种手段,但其行为表现为建立与外部可疑服务器的通信. 因此如何从溯源图中抽象出行为,设定对应的检测和调查规则,是未来可能的技术路线.

(2)基于深度表示学习的攻击调查方法技术路线展望

为解决网安领域标记数据稀缺的问题,未来研究可以尝试引入半监督学习方法. 如生成对抗网络(GAN)、自编码器和扩散模型等. 通过对未标注数据进行自学习或者数据增强,利用有标记数据对半监督生成模型进行参数调整. 基于生成模型得到更多攻击样本,充分利用标注数据和未标注数据,缓解数据不平衡问题,提升深度表示学习模型的检测和调查能力.

(3)基于异常检测的攻击调查方法技术路线展望

现有异常检测方法[126]面临严重的概念漂移问题. 增量学习使得异常检测模型能够适应正常行为模式的变化,能够缓解概念漂移问题. 通过定期将最新的正常行为数据纳入训练集,模型可以逐步更新其对正常行为的理解. 采用滑动窗口等技术来选择用于增量学习的数据,定期更新模型以反映最近的行为模式. 持续评估模型在新数据上的性能,必要时进行模型微调或重训练,以确保异常检测模型性能.

针对基于因果分析、深度表示学习和异常检测的攻击调查方法所面临的问题挑战,本文提出了一系列可能的技术路线. 除了本文总结的三类方法,越来越多新技术的发展和引入能够带来新的解决方案,例如利用向量数据库存储压缩溯源图,提高威胁狩猎与攻击调查的检索效率;或是利用大模型能力,自动分析多源异构日志(如安全设备日志,应用程序日志,网络流量日志等),提取关键字段信息,整合来自不同数据源的信息,从而更全面、准确地描述攻击事件.

8.2　未来研究思路展望

随着信息技术迅速发展,网络安全对抗日益激烈,我国网络安全事件层出不穷,网络空间局势十分严峻. 许多APT组织利用网络攻击工具,在入侵我国重要机构后长期潜伏,这些工具功能强大、结构复杂、隐蔽性高,难以发现. 而基于溯源图的攻击调查能够对攻击场景进行细粒度分析与重建,有助于提升安全防御水平和攻击认知水平. 通过对现有基于溯源图的攻击调查方法进行分类、总结,详细分析部分方法的实验结果,本节提出几个未来值得关注的研究思路方向.

(1)开源威胁情报利用

现有攻击调查方法主要依赖攻击者行为特征和威胁线索,对网络原始日志、终端日志等各种数据进行积极主动搜寻,以还原安全风险与威胁的全貌. 通过收集利用开源威胁情报[127],能够低成本获取多样化、场景化的攻击行为. 该类方法需要突破的关键技术是非结构化文本威胁情报的信息抽取以及语义降级,通过将非结构化文本情报知识转化为单步精确、多步关联的攻击场景图,能够指导攻击调查任务的方向,提升调查效率.

(2)外部知识引入

网络攻击的激增,导致现有攻击调查方法疲于应对,研究人员提出引入外部安全知识库的方法缓解安全场景分析的大量人力物力需求,以MITRE旗下的

ATT&CK 框架应用为例：该框架对攻击者使用的技术和策略进行分类和归纳，我们可以将其形式化为溯源图中的攻击路径，从而更好地了解攻击者的意图．通过理解 ATT&CK 框架下 TTPs，自动生成战术模板图．针对真实网络攻击生成的溯源图，对复杂攻击按战术模板进行拆解，观察各阶段网络攻击情况，使得网络安全分析师拥有高层次的攻击场景视角，采取必要的行动阻止或缓解网络攻击带来的影响．

(3)高质量数据集与实验标准建立

基于溯源图的攻击调查领域缺乏高质量、统一化的数据集，DARPA TC 数据集作为最常用的数据集，其中的数据格式异构，解析困难，且对攻击事件的标注依赖于复杂的、非结构化的攻击报告，数据集解析、标注成本高，所以构建易于解析、标注完备的攻击场景数据集迫在眉睫．当前基于溯源图的攻击调查相关实验中，缺乏统一的实验规范和评价指标，使得不同方法的横向比较十分困难，所以建立统一数据集、统一实验标准，能够方便各类方法进行对比，深度挖掘攻击场景语义，推动领域健康发展．

(4)轻量化人工智能模型应用

人工智能技术正在进一步赋能网络安全与攻击调查相关研究，然而网络攻击调查与响应有较高的实时性要求．未来的轻量化模型将更加注重在不损失关键信息的前提下，通过模型优化和压缩技术减小模型的体积和计算资源需求．加快模型推理速度，提高其在实时攻击调查中的适用性．或是通过整合轻量化人工智能模型与自动化的攻击调查工作流程，包括自动化的报告生成、警报响应和决策支持系统等，提高整体调查效率．总的来说，未来轻量化人工智能模型在基于溯源图的攻击调查任务中的发展方向将更加侧重于提高模型的适用性、灵活性和通用性，以更好地应对不断演变的网络安全挑战．

9 总结

本文对基于溯源图的网络攻击调查方法进行了综述和梳理．针对现代网络攻击的复杂性和隐蔽性，攻击调查成为重要应对方式，通过以警报点为线索进行前后向分析，还原完整攻击路径．然而，传统手动调查流程缓慢且应急响应所需时间长．为了解决这一问题，引入溯源图使得调查过程能够在更精细的审计日志粒度上自动进行．本文总结了近年来攻击调查领域的研究进展，提出核心需求，并将基于溯源图的攻击调查方法分为三类：基于因果分析、基于深度表示学习和基于异常检测的方法．每一类方法都具有独特的流程框架和关键思想．本文从实际应用的角度出发，分析了溯源图优化方法在推动基于溯源图的攻击调查系统产业落地过程中所作出的巨大贡献．为了比较各类方法的效果，本文选择了代表性的方法进行横向实验效果对比．通过对方法和实验效果的综合分析，我们提出了一些可能解决问题的技术路线以及值得关注的未来研究方向．本文为未来的攻击调查研究提供了一些有价值的参考和指导．在进一步探索和改进基于溯源图的攻击调查方法的同时，我们期待这些研究能够为应对复杂网络攻击提供更高效的解决方案．

参考文献

[1] Trellix. Stuxnet: What is Stuxnet? [EB/OL]. (2023) [2023]. https://www.mcafee.com/enterprise/en-hk/security-awareness/ransomware/what-is-stuxnet.html.

[2] 付钰, 李洪成, 吴晓平, 等．基于大数据分析的 APT 攻击检测研究综述[J]. 通信学报, 2015, 36(11): 1-14.
FU Y, LI H C, WU X P, et al. Detecting APT attacks: A survey from the perspective of big data analysis[J]. Journal on Communications, 2015, 36(11): 1-14. (in Chinese)

[3] 吕广旭．基于机器学习的 APT 攻击流量异常检测方法研究[D]. 廊坊: 防灾科技学院, 2023.
LV G X. Research on the Method of APT Attack Traffic Anomaly Detection Based on Machine Learning[D]. Langfang: Institute of Disaster Prevention, 2023. (in Chinese)

[4] 陈泽红．基于自适应模糊聚类的无监督 APT 攻击检测方法研究[J]. 网络安全技术与应用, 2023(7): 45-47.
CHEN Z H. Research on unsupervised APT attack detection method based on adaptive fuzzy clustering[J]. Network Security Technology & Application, 2023(7): 45-47. (in Chinese)

[5] MANDIANT. MANDIANT: Exposing one of China's Cyber Espionage Units[EB/OL]. (2016-03-13) [2023-04-13]. https://vipread.com/library/item/393.

[6] MAO B F, LIU J, LAI Y X, et al. MIF: A multi-step attack scenario reconstruction and attack chains extraction method based on multi-information fusion[J]. Computer Networks, 2021, 198: 108340.

[7] ZHANG X, WU T, ZHENG Q H, et al. Multi-step attack detection based on pre-trained hidden Markov models[J]. Sensors, 2022, 22(8): 2874.

[8] LI Y Z, LI Y M, WU B Y, et al. Invisible backdoor attack with sample-specific triggers[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2021: 16443-16452.

[9] LI S F, XUE M H, ZHAO B Z H, et al. Invisible backdoor attacks on deep neural networks via steganography and regularization[J]. IEEE Transactions on Dependable and Se-

cure Computing, 2021, 18(5): 2088-2105.

[10] MILAJERDI S M, ESHETE B, GJOMEMO R, et al. Propatrol: Attack investigation via extracted high-level tasks[C]//International Conference on Information Systems Security. Cham: Springer, 2018: 107-126.

[11] KING S T, CHEN P M. Backtracking intrusions[C]//Proceedings of the nineteenth ACM symposium on Operating systems principles. New York: ACM, 2003: 223-236.

[12] TAN C, WANG Q, WANG L N, et al. Attack provenance tracing in cyberspace: Solutions, challenges and future directions[J]. IEEE Network, 2019, 33(2): 174-180.

[13] 冷涛, 蔡利君, 于爱民, 等. 基于系统溯源图的威胁发现与取证分析综述[J]. 通信学报, 2022, 43(7): 172-188.
LENG T, CAI L J, YU A M, et al. Review of threat discovery and forensic analysis based on system provenance graph[J]. Journal on Communications, 2022, 43(7): 172-188. (in Chinese)

[14] LI Z Y, CHEN Q A, YANG R Q, et al. Threat detection and investigation with system-level provenance graphs: A survey [EB/OL]. (2020-06-02) [2023-04-13]. http://arxiv.org/abs/2006.01722.

[15] INAM M A, CHEN Y F, GOYAL A, et al. SoK: History is a vast early warning system: Auditing the provenance of system intrusions[C]//2023 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE, 2023: 2620-2638.

[16] 潘亚峰, 朱俊虎, 周天阳. APT攻击场景重构方法综述[J]. 信息工程大学学报, 2021, 22(1): 55-60, 80.
PAN Y F, ZHU J H, ZHOU T Y. Survey on APT attack scenario reconstruction methods[J]. Journal of Information Engineering University, 2021, 22(1): 55-60, 80. (in Chinese)

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

[18] HAN X Y, PASQUIER T, SELTZER M. Provenance-based intrusion detection: Opportunities and challenges [C]//Proceedings of the 10th USENIX Conference on Theory and Practice of Provenance. Berkeley: USENIX Association, 2018: 3.

[19] 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.

[20] RAFAILIDIS D, AXENOPOULOS A, ETZOLD J, et al. Content-based tag propagation and tensor factorization for personalized item recommendation based on social tagging[J]. ACM Transactions on Interactive Intelligent Systems, 2014, 3(4): 26.

[21] 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.

[22] KURNIAWAN K, EKELHART A, KIESLING E, et al. KRYSTAL: Knowledge graph-based framework for tactical attack discovery in audit data[J]. Computers & Security, 2022, 121: 102828.

[23] XIONG C L, ZHU T T, DONG W H, et al. Conan: A practical real-time APT detection system with high accuracy and efficiency[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19(1): 551-565.

[24] ZHU T T, YU J K, CHEN T M, et al. APTSHIELD: A stable, efficient and real-time APT detection system for linux hosts[EB/OL]. (2021-12-16) [2023-05-03]. http://arxiv.org/abs/2112.09008.

[25] LEE K H, ZHANG X Y, XU D Y. High accuracy attack provenance via binary-based execution partition[C]//20th Annual Network and Distributed System Security Symposium. San Diego: Internet Society, 2013: 1-16.

[26] 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. San Diego: Internet Society, 2016: 1-15.

[27] MA S Q, Z J, W F, et al. MPI: Multiple perspective attack investigation with semantic aware execution partitioning [C]//Proceedings of the 26th USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2017: 1111-1128.

[28] ALHANAHNAH M, MA S Q, GEHANI A, et al. AutoMPI: Automated multiple perspective attack investigation with semantics aware execution partitioning[J]. IEEE Transactions on Software Engineering, 2023, 49(4): 2761-2775.

[29] D´ELIA D C, COPPA E, NICCHI S, et al. SoK: Using dynamic binary instrumentation for security (and how you may get caught red handed) [C]//Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security. New York: ACM, 2019: 15-27.

[30] 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. San Diego: Internet Society, 2020: 24270.

[31] PAN Y, GE X T, FANG C R, et al. A systematic literature review of android malware detection using static analysis[J]. IEEE Access, 2020, 8: 116363-116379.

[32] AGHAKHANI H, GRITTI F, MECCA F, et al. When malware is packin' heat; limits of machine learning classifiers based on static analysis features[C]//Proceedings 2020 Network and Distributed System Security Symposium. Reston: Internet Society, 2020: 24310.

[33] MOSSBERG M, MANZANO F, HENNENFENT E, et al. Manticore: A user-friendly symbolic execution framework for binaries and smart contracts[C]//2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE). Piscataway: IEEE, 2019: 1186-1189.

[34] HE J X, BALUNOVIĆ M, AMBROLADZE N, et al. Learning to fuzz from symbolic execution with application to smart contracts[C]//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2019: 531-548.

[35] POEPLAU S, FRANCILLON A. Symbolic execution with SymCC: Don't interpret, compile! [C]//Proceedings of the 29th USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2020: 181-198.

[36] 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. Reston: Internet Society, 2021: 1-18.

[37] MUGGLETON S H, LIN D H, TAMADDONI-NEZHAD A. Meta-interpretive learning of higher-order dyadic datalog: Predicate invention revisited[J]. Machine Learning, 2015, 100(1): 49-73.

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

[39] KWON Y, KIM D, SUMNER W N, et al. LDX: Causality inference by lightweight dual execution[C]//Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems. New York: ACM, 2016: 503-515.

[40] YANG R, MA S, XU H, et al. UIScope: Accurate, instrumentation-free, and visible attack investigation for GUI applications[C]//27th Annual Network and Distributed System Security Symposium. Reston: Internet Society, 2020: 1-18.

[41] ALBERT B, TULLIS T. Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics [M]. 3rd ed. Amsterdam: Morgan Kaufmann, 2022.

[42] MILAJERDI S M, GJOMEMO R, ESHETE B, et al. HOLMES: Real-time APT detection through correlation of suspicious information flows[EB/OL]. (2018-10-03) [2023-05-03]. http://arxiv.org/abs/1810.01594.

[43] JI S X, PAN S R, CAMBRIA E, et al. A survey on knowledge graphs: Representation, acquisition, and applications [J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 33(2): 494-514.

[44] KOSTYLEV E V, REUTTER J L, ROMERO M, et al. SPARQL with property paths[C]//International Semantic Web Conference. Cham: Springer, 2015: 3-18.

[45] 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.

[46] HANNOUSSE A, YAHIOUCHE S. Handling webshell attacks: A systematic mapping and survey[J]. Computers & Security, 2021, 108: 102366.

[47] CHAI Y H, DU L, QIU J, et al. Dynamic prototype network based on sample adaptation for few-shot malware detection[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(5): 4754-4766.

[48] CHAI Y H, QIU J, YIN L H, et al. From data and model levels: Improve the performance of few-shot malware classification[J]. IEEE Transactions on Network and Service Management, 2022, 19(4): 4248-4261.

[49] SYAKUR M A, KHOTIMAH B K, ROCHMAN E S, et al. Integration K-means clustering method and elbow method for identification of the best customer profile cluster[J]. IOP Conference Series: Materials Science and Engineering, 2018, 336: 012017.

[50] JELODAR H, WANG Y L, YUAN C, et al. Latent dirichlet allocation (LDA) and topic modeling: Models, applications, a survey[J]. Multimedia Tools and Applications, 2019, 78(11): 15169-15211.

[51] KHAZAEI A, GHASEMZADEH M, DERHAMI V. An automatic method for CVSS score prediction using vulnerabilities description[J]. Journal of Intelligent & Fuzzy

Systems, 2015, 30(1): 89-96.

[52] YADAV T, RAO A M. Technical aspects of cyber kill chain[C]//International Symposium on Security in Computing and Communication. Cham: Springer, 2015: 438-452.

[53] ZOU H T, GONG Z G, ZHANG N, et al. TrustRank: A Cold-Start tolerant recommender system[J]. Enterprise Information Systems, 2015, 9(2): 117-138.

[54] TIAN Z H, SHI W, TAN Z Y, et al. Deep learning and dempster-shafer theory based insider threat detection [J/OL]. Mobile Networks and Applications, 2020. https://doi.org/10.1007/s11036-020-01656-7.

[55] ALLEY E C, KHIMULYA G, BISWAS S, et al. Unified rational protein engineering with sequence-based deep representation learning[J]. Nature Methods, 2019, 16: 1315-1322.

[56] ZHU A Z, YUAN L Z, CHANEY K, et al. Unsupervised event-based learning of optical flow, depth, and egomotion[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 989-997.

[57] HIGUERA J R B, HIGUERA J B, GARCÍA J L T, et al. Building a dataset through attack pattern modeling and analysis system[J]. Computers & Electrical Engineering, 2022, 97: 107614.

[58] PRABAKARAN S, RAMAR R, HUSSAIN I, et al. Predicting attack pattern via machine learning by exploiting stateful firewall as virtual network function in an SDN network[J]. Sensors, 2022, 22(3): 709.

[59] 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.

[60] SHARMA S, GOSAIN A, JAIN S. A review of the oversampling techniques in class imbalance problem[C]//International Conference on Innovative Computing and Communications. Singapore: Springer, 2022: 459-472.

[61] MOHAMMED R, RAWASHDEH J, ABDULLAH M. Machine learning with oversampling and undersampling techniques: Overview study and experimental results[C]// 2020 11th International Conference on Information and Communication Systems (ICICS). Piscataway: IEEE, 2020: 243-248.

[62] BAHADIR C D, WANG A Q, DALCA A V, et al. Deep-learning-based optimization of the under-sampling pattern in MRI[J]. IEEE Transactions on Computational Imaging, 2020, 6: 1139-1152.

[63] DAI Q, LIU J W, LIU Y. Multi-granularity relabeled under-sampling algorithm for imbalanced data[J]. Applied Soft Computing, 2022, 124: 109083.

[64] YU Y, SI X S, HU C H, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural Computation, 2019, 31(7): 1235-1270.

[65] 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.

[66] ZHOU Z H. Semi-supervised learning[M]//Machine Learning. Singapore: Springer, 2021: 315-341.

[67] 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.

[68] 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). Piscataway: IEEE, 2022: 540-557.

[69] XIA F, LIU J Y, NIE H S, et al. Random walks: A review of algorithms and applications[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2020, 4(2): 95-107.

[70] LAZARIDOU A, PHAM N T, BARONI M. Combining language and vision with a multimodal skip-gram model [EB/OL]. (2015-06-12)[2023-05-03]. http://arxiv.org/abs/1501.02598.

[71] CIPRESSO P, GIGLIOLI I A C, RAYA M A, et al. The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature [J]. Frontiers in Psychology, 2018, 9: 2086.

[72] CHUNAEV P. Community detection in node-attributed social networks: A survey[J]. Computer Science Review, 2020, 37: 100286.

[73] TENG X Y, LIU J, LI M M. Overlapping community detection in directed and undirected attributed networks using a multiobjective evolutionary algorithm[J]. IEEE Transactions on Cybernetics, 2021, 51(1): 138-150.

[74] ALSENTZER E, FINLAYSON S G, LI M M, et al. Subgraph neural networks[C]//Proceedings of the 34th International Conference on Neural Information Processing Systems. New York: ACM, 2020: 8017-8029.

[75] RIBEIRO P, PAREDES P, SILVA M E P, et al. A survey on subgraph counting: Concepts, algorithms, and applications to network motifs and graphlets[J]. ACM Computing Surveys, 54(2): 28.

[76] HU W, ARDESHIRICHAM A, KASTNER R. Hardware information flow tracking[J]. ACM Computing Surveys, 54(4): 83.

[77] LALITHSENA S, PERERA S, KAPANIPATHI P, et al. Domain-specific hierarchical subgraph extraction: A recommendation use case[C]//2017 IEEE International Conference on Big Data (Big Data). Piscataway: IEEE, 2017: 666-675.

[78] SCHUMAN C D, HAMILTON K, MINTZ T, et al. Shortest path and neighborhood subgraph extraction on a spiking memristive neuromorphic implementation[C]//Proceedings of the 7th Annual Neuro-inspired Computational Elements Workshop. New York: ACM, 2019: 1-6.

[79] CAI H Y, ZHENG V W, CHANG K C C. A comprehensive survey of graph embedding: Problems, techniques, and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(9): 1616-1637.

[80] 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. Reston: Internet Society, 2021: 1-18.

[81] TARJAN R. Depth-first search and linear graph algorithms[J]. SIAM Journal on Computing, 1972, 1(2): 146-160.

[82] BUTTCHER S, CLARKE C L, CORMACK G V. Information Retrieval: Implementing and Evaluating Search Engines[M]. Cambridge: MIT Press, 2016.

[83] SEN P C, HAJRA M, GHOSH M. Supervised classification algorithms in machine learning: A survey and review [C]//Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018. Singapore: Springer, 2020: 99-111.

[84] YIM O, RAMDEEN K T. Hierarchical cluster analysis: Comparison of three linkage measures and application to psychological data[J]. The Quantitative Methods for Psychology, 2015, 11(1): 8-21.

[85] THONGTAN T, PHIENTHRAKUL T. Sentiment classification using document embeddings trained with cosine similarity[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. Stroudsburg: ACL, 2019: 407-414.

[86] AGRAWAL S, AGRAWAL J. Survey on anomaly detection using data mining techniques[J]. Procedia Computer Science, 2015, 60: 708-713.

[87] 李忠, 靳小龙, 庄传志, 等. 面向图的异常检测研究综述[J]. 软件学报, 2021, 32(1): 167-193.
LI Z, JIN X L, ZHUANG C Z, et al. Overview on graph based anomaly detection[J]. Journal of Software, 2021, 32(1): 167-193. (in Chinese)

[88] HASSAN W U, GUO S, LI D, et al. Nodoze: Combatting threat alert fatigue with automated provenance triage[C]//26th Annual Network and Distributed System Security Symposium. Reston: Internet Society, 2019: 1-15.

[89] 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.

[90] 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.

[91] LIU F, WEN Y, ZHANG D, et al. Log2vec: A heterogeneous graph embedding based approach for detecting cyber threats within enterprise[C]//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2019: 1777-1794.

[92] 石川, 王睿嘉, 王啸. 异质信息网络分析与应用综述[J]. 软件学报, 2022, 33(2): 598-621.
SHI C, WANG R J, WANG X. Survey on heterogeneous information networks analysis and applications[J]. Journal of Software, 2022, 33(2): 598-621. (in Chinese)

[93] HOU L, LI J M, GU Z Q, et al. PANNER: POS-aware nested named entity recognition through heterogeneous graph neural network[J/OL]. IEEE Transactions on Computational Social Systems, 2022. https://ieeexplore. ieee. org/document/9745261.

[94] LI Z T, CHENG X, SUN L X, et al. A hierarchical approach for advanced persistent threat detection with attention-based graph neural networks[J]. Security and Communication Networks, 2021, 2021: 9961342.

[95] ZENGY J, WANG X, LIU J, et al. SHADEWATCHER: Recommendation-guided cyber threat analysis using system audit records[C]//2022 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE, 2022: 489-506.

[96] MANZOOR E, MILAJERDI S M, AKOGLU L. Fast memory-efficient anomaly detection in streaming heterogeneous graphs[C]//Proceedings of the 22nd ACM SIG-

KDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1035-1044.

[97] HAN X Y, PASQUIER T, BATES A, et al. UNICORN: Runtime provenance-based detector for advanced persistent threats[C]//Network and Distributed Systems Security (NDSS) Symposium 2020. Reston: Internet Society, 2020: 24046.

[98] RIECK B, BOCK C, BORGWARDT K. A persistent weisfeiler-lehman procedure for graph classification[C]//Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019: 5448-5458.

[99] 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.

[100] CHURCH K W. Word2Vec[J]. Natural Language Engineering, 2017, 23(1): 155-162.

[101] DOUZI S, AMAR M, OUAHIDI B EL, et al. Towards a new spam filter based on PV-DM (paragraph vector-distributed memory approach) [J]. Procedia Computer Science, 2017, 110: 486-491.

[102] WANG S, WANG Z L, ZHOU T, et al. ThreaTrace: Detecting and tracing host-based threats in node level through provenance graph learning[EB/OL]. (2021-11-08)[2023-05-03]. http://arxiv.org/abs/2111.04333.

[103] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 1025-1035.

[104] BORDES A, USUNIER N, GARCIA-DURÁN A, et al. Translating embeddings for modeling multi-relational data[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. New York: ACM, 2013: 2787-2795.

[105] LIN Y K, LIU Z Y, SUN M S, et al. Learning entity and relation embeddings for knowledge graph completion[C]// Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. New York: ACM, 2015: 2181-2187.

[106] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010.

[107] HAN K, XIAO A, WU E H, et al. Transformer in transformer[C]//Advances in Neural Information Processing Systems. Virtual: PMLR, 2021: 15908-15919.

[108] YANG D Q, LI B, RETTIG L, et al. HistoSketch: Fast similarity-preserving sketching of streaming histograms with concept drift[C]//2017 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE, 2017: 545-554.

[109] NARAYANAN A, CHANDRAMOHAN M, VENKATESAN R, et al. Graph2vec: Learning distributed representations of graphs[EB/OL]. (2017-07-17) [2023-05-03]. http://arxiv.org/abs/1707.05005.

[110] MA M X, NGAN H Y T, LIU W. Density-based outlier detection by local outlier factor on largescale traffic data [J]. Electronic Imaging, 2016, 28(14): 1-4.

[111] PARK H S, JUN C H. A simple and fast algorithm for K-medoids clustering[J]. Expert Systems with Applications, 2009, 36(2): 3336-3341.

[112] SHANI G, GUNAWARDANA A. Evaluating recommendation systems[M]//Recommender Systems Handbook. Boston: Springer, 2011: 257-297.

[113] GUO Z X, ZHU L G, HAN L. Research on short text classification based on RoBERTa-TextRCNN[C]//2021 International Conference on Computer Information Science and Artificial Intelligence (CISAI). Piscataway: IEEE, 2021: 845-849.

[114] LEE K H, ZHANG X Y, XU D Y. LogGC: Garbage collecting audit log[C]//Proceedings of the 2013 ACM SIGSAC conference on Computer & Communications Security. New York: ACM, 2013: 1005-1016.

[115] TANG Y T, LI D, LI Z C, et al. NodeMerge: Template based efficient data reduction for big-data causality analysis[C]//Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2018: 1324-1337.

[116] XU Z, WU Z Y, LI Z C, et al. High fidelity data reduction for big data security dependency analyses[C]//Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM, 2016: 504-516.

[117] HOSSAIN M N, WANG J, WEISSE O, et al. Dependence-preserving data compaction for scalable forensic analysis[C]//Proceedings of the 27th USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2018: 1723-1740.

[118] ZHU T T, WANG J Y, RUAN L Q, et al. General, efficient, and real-time data compaction strategy for APT forensic analysis[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 3312-3325.
[119] FEI P, LI Z, WANG Z, et al. SEAL: Storage-efficient causality analysis on enterprise logs with query-friendly compression[C]//Proceedings of the 30th USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2021: 2987-3004.
[120] DING H, YAN S, ZHAI J, et al. ELISE: A storage efficient logging system powered by redundancy reduction and representation learning[C]///Proceedings of the 30th USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2021: 3023-3040.
[121] DING H L, ZHAI J, DENG D, et al. The case for learned provenance graph storage systems[C]//Proceedings of the 32nd USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2023: 3277-3294.
[122] ANJUM M M, IQBAL S, HAMELIN B. Analyzing the Usefulness of the DARPA OpTC dataset in cyber threat detection research[C]//Proceedings of the 26th ACM Symposium on Access Control Models and Technologies. New York: ACM, 2021: 27-32.
[123] HASSAN W U, BATES A, MARINO D. Tactical provenance analysis for endpoint detection and response systems[C]//2020 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE, 2020: 1172-1189.
[124] DONG F, WANG L, NIE X, et al. DISTDET: A cost-effective distributed cyber threat detection system[C]//Proceedings of the 32nd USENIX Conference on Security Symposium. Berkeley: USENIX Association, 2023: 6575-6592.
[125] 轩勃娜, 李进. 基于改进CNN的恶意软件分类方法[J]. 电子学报, 2023, 51(5): 1187-1197.
XUAN B N, LI J. Malware classification method based on improved CNN[J]. Acta Electronica Sinica, 2023, 51(5): 1187-1197. (in Chinese)
[126] 严莉, 张凯, 徐浩, 等. 基于图注意力机制和Transformer的异常检测[J]. 电子学报, 2022, 50(4): 900-908.
YAN L, ZHANG K, XU H, et al. Abnormal detection based on graph attention mechanisms and Transformer[J]. Acta Electronica Sinica, 2022, 50(4): 900-908. (in Chinese)
[127] 郑锐, 汪秋云, 林卓庞, 等. 一种基于威胁情报层次特征集成的挖矿恶意软件检测方法[J]. 电子学报, 2022, 50(11): 2707-2715.
ZHENG R, WANG Q Y, LIN Z P, et al. Cryptojacking malware hunting: A method based on ensemble learning of hierarchical threat intelligence feature[J]. Acta Electronica Sinica, 2022, 50(11): 2707-2715. (in Chinese)

作者简介

仇 晶 女,博士. 广州大学网络空间安全学院教授,博士生导师. 主要研究领域为网络安全、人工智能及大数据安全. 中国电子学会会员编号:E190035636M.

陈荣融 男,广州大学网络空间安全学院硕士研究生. 主要研究领域为网络攻击检测与调查.

朱浩瑾 男,博士. 上海交通大学计算机科学与工程系教授,博士生导师,国家杰出科学基金获得者,IEEE Fellow,上海交通大学电子信息与电气工程学院副院长. 主要研究领域为物联网安全.

肖岩军 男,学士. 绿盟科技平行实验室主任研究员. 主要研究领域为态势感知、知识图谱、APT追踪、人工智能决策指挥、网络靶场.

殷丽华 女,博士. 广州大学网络空间安全学院教授,博士生导师,国家高层次人才. 主要研究领域为网络安全、物联网安全、大数据安全与隐私保护.

田志宏 男,博士. 广州大学党委常委,副校长,广州大学网络空间安全学院教授,博士生导师. 主要研究领域为网络攻防对抗、漏洞挖掘与利用、APT检测与溯源、工控安全.
E-mail: tianzhihong@gzhu.edu.cn