基于APT活动全生命周期的攻击与检测综述

王郅伟[1,2]，何睎杰[1,2]，易鑫[1]，李孜旸[1,2]，曹旭栋[1,2]，尹涛[2]，李书豪[2]，付安民[3]，张玉清[1,2,4]
（1.中国科学院大学国家计算机网络入侵防范中心，北京 101408；2.中关村实验室，北京 100194；
3. 南京理工大学计算机科学与工程学院，江苏 南京 210094；4.海南大学网络空间安全学院，海南 海口 570228）

摘　要：从攻击方法和检测方法两方面展开，首先综述高级持续威胁（APT）攻击的定义与特点，总结相关攻击模型的研究发展，在此基础上给出更一般性的APT全生命周期模型，并划分4个阶段，信息收集阶段、入侵实施阶段、内网攻击阶段和数据渗出阶段，对每一个阶段，重点调研近5年的研究论文，归纳总结各阶段的攻击与检测技术，并给出分析。最后，结合APT攻防技术相互博弈、快速发展的趋势，指出了当前攻防双方面临的挑战和未来研究的发展方向。

关键词：高级持续威胁；网络杀伤链模型；全生命周期；零日攻击；检测
中图分类号：TP399
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2024128

Survey of attack and detection based on the full life cycle of APT

WANG Zhiwei[1,2], HE Xijie[1,2], YI Xin[1], LI Ziyang[1,2], CAO Xudong[1,2], YIN Tao[2], LI Shuhao[2], FU Anmin[3], ZHANG Yuqing[1,2,4]
1. National Computer Network Intrusion Protection Center, University of Chinese Academy of Sciences, Beijing 101408, China
2. Zhongguancun Laboratory, Beijing 100194, China
3. School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
4. School of Cyberspace Security (School of Cryptography), Hainan University, Haikou 570228, China

Abstract: The advanced persistent threat (APT) attack was explored from two perspectives: attack methods and detection methods. First, the definitions and characteristics of APT attacks were reviewed and the development of related attack models was summarized. Based on this, a more general APT full lifecycle model was proposed, which was divided into four stages: information gathering, intrusion execution, internal network penetration, and data exfiltration. For each stage, recent research papers from the past five years were thoroughly reviewed, and the attack and detection techniques for each stage were analyzed. Finally, in light of the dynamic landscape of APT attack and defense technologies, the paper underscores the formidable challenges confronting both offense and defense and offers guidance for future research in this domain.

Keywords: APT, network kill chain model, full life cycle, 0day attack, detect

收稿日期：2024–02–18；**修回日期：**2024–05–24
通信作者：张玉清，zhangyq@nipc.org.cn
基金项目：国家重点研发计划基金资助项目（No.2023YFB3106400, No.2023QY1202）；国家自然科学基金资助项目（No.U2336203, No.U1836210）；海南省重点研发计划基金资助项目（No.GHYF2022010）；北京市自然科学基金资助项目（No.4242031）
Foundation Items: The National Key Research and Development Program of China (No.2023YFB3106400, No.2023QY1202), The National Natural Science Foundation of China (No. U2336203, No. U1836210), The Key Research and Development Program of Hainan Province (No.GHYF2022010), The Beijing Natural Science Foundation (No.4242031)

0　引言

随着互联网和物联网技术的高速发展，人类的生产、生活、教育和娱乐等行为愈发地与网络高度耦合。在网络空间博弈与国际局势演变的大背景下，具有强组织性、高度技术复杂性、针对性和隐蔽性的高级持续威胁（APT, advanced persistent threat）攻击对不同行业的网络资产威胁日渐加深[1]。以2010年震网（Stuxnet）蠕虫病毒为例，APT攻击者利用高度精准的社会工程学策略，针对伊朗核项目的工作人员进行了攻击。这导致伊朗用于浓缩铀的离心机系统遭受严重破坏，从而迟滞了伊朗核项目的进展。2023年6月，卡巴斯基揭露的一起APT攻击事件[1]提到有APT组织利用了iOS系统中多个零日（0day）漏洞组合，使用IMessage信息服务的0-Click-0day在全球范围进行大规模攻击。从该攻击活动的目标范围、复杂度、攻击技术和跨越时间来看，这是近10年内最顶尖的国家级APT攻击活动[2]。攻击者展示的高超技术和丰富资源，凸显了APT攻击威胁的不断加剧，对APT攻击模式的厘清、对APT攻击的防御与检测也刻不容缓。

本文主要对中外期刊论文、EI数据库、CCF推荐网络与信息安全国际学术顶级会议（如IEEE S&P、USENIX Security、ACM CCS、NDSS）中发表的APT攻击与检测相关论文进行了深入调研分析，具体的年份与类型分布如图1所示，相关研究集中在近5年，5年之前的文献频次较低且平均水平稳定，之前的研究侧重于APT攻击的梳理，而近5年则更专注于APT检测技术的发展，体现了检测技术相较于攻击技术相对滞后的特点。本文仅选择具有突出意义的研究进行介绍，同时，与APT相关的研究工作一直是学术研究中的重点，对传统技术的总结，对新型研究的介绍也具有重要的现实意义。

与现有综述相比，本文不仅仅关注某一领域或某一类分析方法的APT攻击研究[3-5]，亦不局限于单一的APT攻击抑或APT检测方面[6-7]，本文研究纵深结合两者，更不仅对APT本身概念和特征进行探讨[8-10]，本文辅以案例，以及攻击与检测技术中的具体细分方向的研究迭代和最新进展，以APT全生命周期作为核心并展开，全面展现了APT研究发展的动态过程。

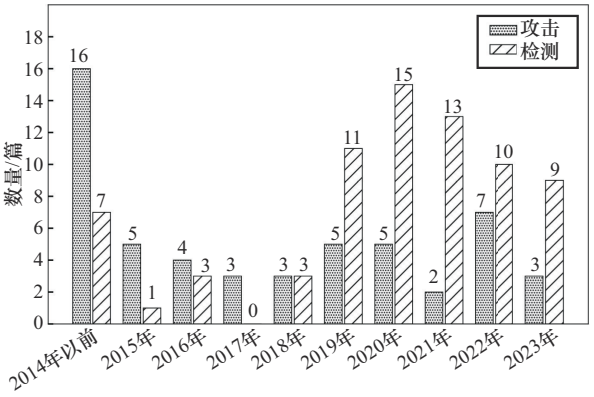


图1　APT文献统计

本文根据APT活动的典型全生命周期结构，从攻击和检测2个角度，重点对近5年的APT相关研究工作进行了分析，主要贡献包括4个方面。

1) 在充分分析APT攻击定义、特点、模型的基础上，整合研究机构对APT攻击模型的总结，提出了更加全面、系统化、科学的APT全生命周期模型，并创新性地以全生命周期阶段作为攻击与检测技术划分，分析和归类APT攻击方法和检测手段，更加清晰直观。

2) 在攻击方面，结合学术研究与真实案例，从APT攻击者的视角，辅以真实APT攻击案例，还原各过程的具体技术，并分析最新研究成果。将APT攻击范畴不仅仅局限于Web攻击方面，随着技术发展的趋势，拓宽到新兴发展的物联网、AI等领域。

3) 在检测方面，以全生命周期为脉络，总结并梳理研究文献中的APT检测技术，对类似的检测方案进行分类分析，并按照时间线梳理使用溯源图神经网络进行全生命周期的检测技术的进展，在此基础上，考察了基于情报共享的全生命周期检测技术落脚于工业界的真实案例。

4) 依据现有技术的发展和APT攻击与检测技术的现状，揭示了APT攻击与检测领域发展的挑战与机遇，并分别从APT攻击和检测的角度分析了未来发展的侧重点。

1　基于APT活动全生命周期的攻击

1.1　APT的定义

APT发展历程如图2所示，普遍认为，APT这一概念最早由美国空军上校Gregory Rattray创造和引用[11]。目前针对APT攻击并未有一个较为权威的定义，普遍认可的描述由美国国家标准与技术研

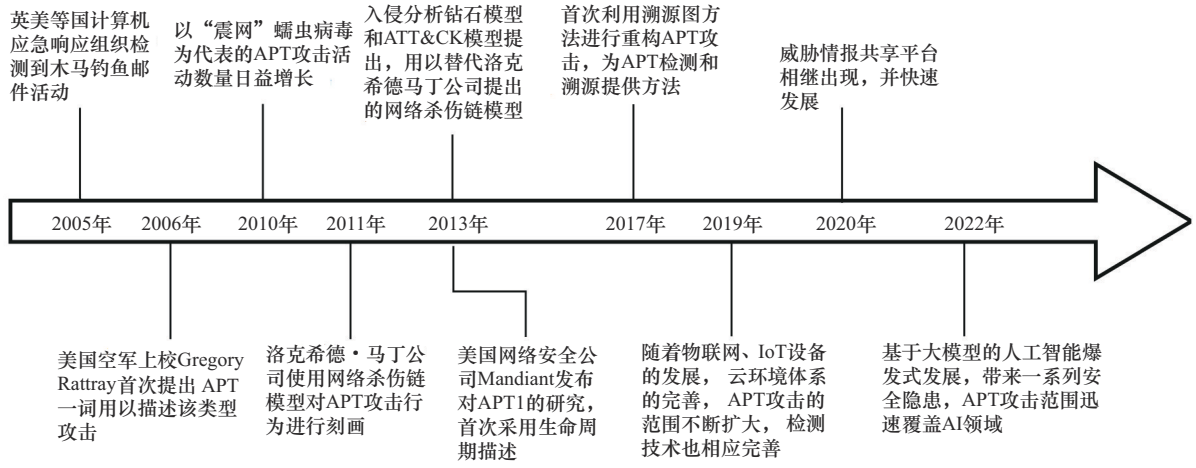


图2 APT发展历程

究院（NIST, National Institute of Standards and Technology）给出：攻击者掌握先进的专业知识和有效的资源，通过多种攻击途径，在特定组织的信息技术基础设施建立并转移立足点，窃取机密信息，破坏或阻碍任务、程序或组织的关键系统，或者驻留组织的内部网络，进行后续攻击[12]。该定义很好地描述了攻击者的技术与资源，攻击目的和手段等信息。然而，随着网络技术的发展，以及APT攻击手段的进化，该描述难以兼顾现代APT攻击的高度针对性以及高度隐蔽性的特点，为此，付钰等[5]在该基础上给出补充定义，并进行总结。本文再进一步补充定义：APT攻击是以攻击者的高水平专业知识和丰富资源为基础，以多种攻击方式为手段，以破坏关键信息基础设施、获取关键敏感信息或阻碍任务实施为目的，以特定组织或信息系统为攻击对象的长期隐蔽网络攻击。

1.2 APT的特点

APT攻击相较于传统网络攻击具有以下6个显著特征：先进专业技术、针对性强、持续时间长、组织严密、威胁程度大和高隐蔽性[2,5]。一般的黑客很难形成具有上述特点的严密组织，因此实施APT攻击的组织通常被认为具有国家级的背景。

从上述特点来看，针对性强是APT攻击区别于传统网络攻击如蠕虫病毒、僵尸网络和勒索病毒的关键特性，而高隐蔽性又是APT攻击与渗透测试的本质区别。

1.3 APT攻击模型

为了更好地研究APT攻击，相关研究机构给出了不同的APT攻击模型，从各种角度，更加全面地分析APT攻击。

Caltagirone等[13]曾提出一种用于分析攻击事件的入侵分析钻石模型，如图3所示。该模型通过4个节点：对手、能力、基础设施和受害者之间的关系，以及相关拓展特征的附加（如拓展的钻石模型，如图4所示）从根本上分析入侵实施的动机、意图和渠道等信息，侧重于从理论和科学层面，指导分析入侵事件。

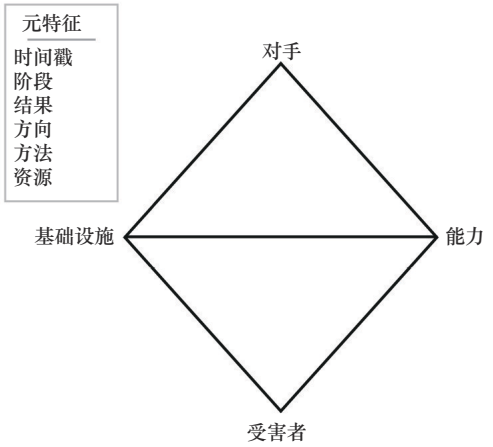


图3 入侵分析钻石模型

具体实施层面，洛克希德·马丁公司曾使用网络杀伤链模型分为7个阶段描述了网络攻击的流程[14]，如图5所示。

网络杀伤链模型起源于网络入侵防御早期，突出了病毒和漏洞相关的外线防守，但是无法完整的涵盖现代APT攻击者攻击手段的复杂性和攻击方式的灵活性，例如社会工程学攻击、供应链投毒[15]。同时，该模型仅对网络入侵进行总结和概

括，缺乏深度和广度的解读，难以形成详尽的系统性知识框架。因此，非营利机构MITRE提出了入侵者战术、技术和共有知识库（ATT&CK, adversarial tactics, techniques, and common knowledge）模型用以替代网络杀伤链模型。

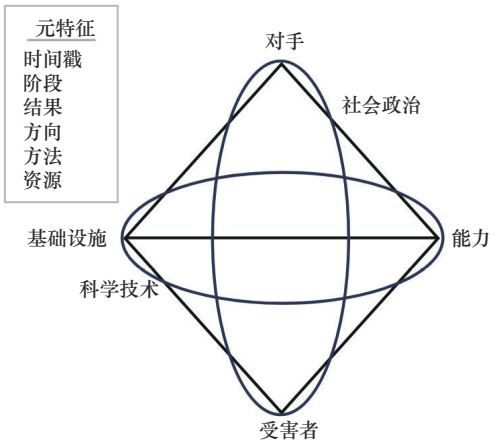


图4 入侵分析钻石模型拓展

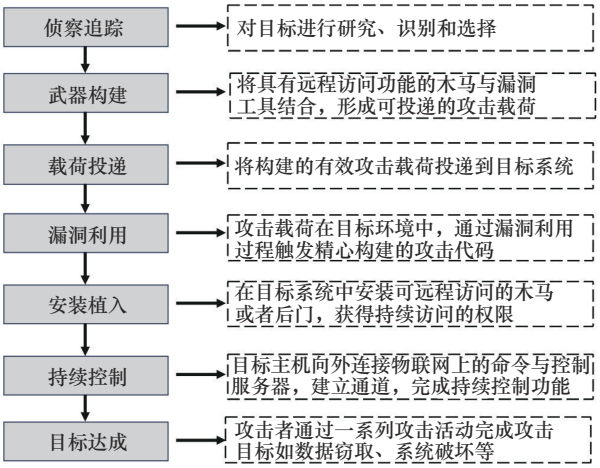


图5　网络杀伤链模型

ATT&CK模型是在网络杀伤链模型的基础上，构建的一套更细粒度，基于共享、威胁情报分析的知识模型和框架。ATT&CK模型将网络攻击划分为14个战术阶段，每个战术阶段可下分若干子技术，涵盖攻击的全部阶段以及所有平台。攻击者可以随意切换战术来实现最终目标，更加细化和灵活。ATT&CK知识库迄今仍然对新型网络攻击技术保持密切追踪，基于ATT&CK矩阵衍生出的攻击与防御方法仍是学术界的研究重点[16]。

在学术研究中，大量的研究人员倾向于使用“全生命周期”刻画APT活动。该概念由美国网络安全公司Mandiant于2013年对APT1的分析报告[17]中首次提出，通过时间或者逻辑上有顺承关系的脉络图，详细地展开APT攻击的描述。Khaleefa等[18]通过侦察、入侵、保持、横向移动和数据渗出5个名词组成的环形结构简单描述了全生命周期的概念，但由于其描述情况过于简单，Talib等[7]的工作对其进行扩展，将APT攻击划分为6个阶段组成的环形，如图6所示，并在此基础上进行讨论。但是图6中的循环体并非真实APT攻击过程，数据泄露阶段完成后并不会重新进行侦察和武器化、传递和初步入侵环节。

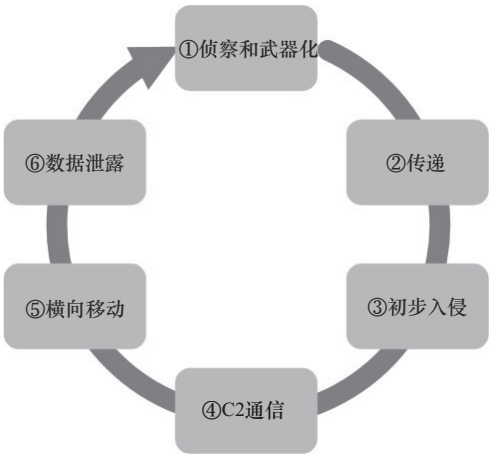


图6　APT攻击的6个阶段

Sharma等[9]则在前人的基础上，较为细致地给出了APT攻击的另一种全生命周期划分，该划分将循环体设置在立足点建立后的攻击流程，使用5个阶段组成的内循环进行标识，更加科学、完整地给出了APT攻击的全生命周期，具体如图7所示。

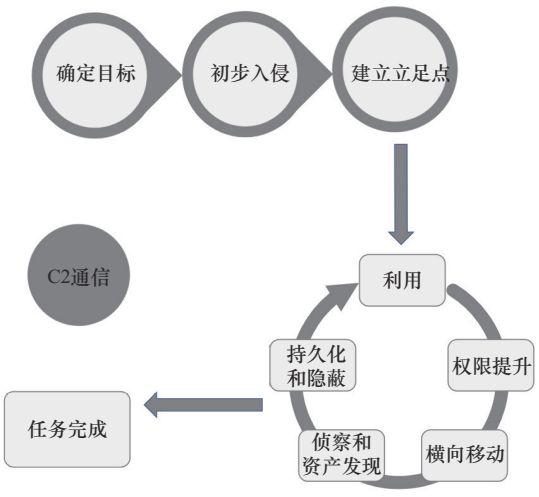


图7　Sharma等[9]提出的APT攻击的全生命周期

由于Sharma等[9]的APT攻击等全生命周期划分过于碎片化，并未形成明确的阶段化过程，导致如命令与控制（C2, command and control）具体与哪部分密切相关并未明确，因此本文参考中国信息安全测评中心[19]2023年4月发布的《全球高级持续性威胁（APT）研究报告》中提出的APT全生命周期，在其基础上提出更科学的全生命周期划分，将攻击过程分为4个战术阶段（信息收集阶段、入侵实施阶段、内网攻击阶段、数据渗出阶段）和6个环节（信息收集、入侵实施、权限提升、横向移动、痕迹清除与持久化和数据渗出），该生命周期如图8所示，本文的分析也将以此划分展开。

值得说明的是，并不是所有APT攻击都会存在数据渗出这一阶段，也有部分APT攻击行为仅为了破坏目标系统，如震网蠕虫病毒。因此上述全生命周期适合一般数据导向的APT攻击，对于非数据导向的APT攻击，仅最后一阶段存在差异。

1.4　APT攻击方法

对上述研究总结得出的APT活动4个阶段进行分析，由于4个阶段的目标、环境不同，不同阶段的攻击特征也存在明显的差异，在对APT攻击方法的分析中，现有的研究也总结了不同阶段所使用的不同攻击方法，以帮助更好地理解APT活动。

1.4.1　信息收集阶段

信息收集阶段是在APT攻击之前进行的侦察活动，攻击者需要了解目标系统的脆弱性，以便制定有效的攻击计划。信息收集阶段可以帮助攻击者收集关于目标的各种信息，攻击者可以利用这些信息来制订更有效的攻击计划，以增加攻击的成功率。

Auty等[10]认为，在APT攻击的信息收集阶段，攻击者所收集的情报信息主要有2类：①目标机构背景、业务范围、职员信息的人文信息；②目标信息系统的网络资产、网站指纹、网站架构以及应用程序版本等信息。前者的信息收集与整理，基于社会工程学方法，为后续针对人员或个体实施鱼叉攻击或水坑攻击做铺垫；后者的信息收集与整理，基于网站脆弱点寻找，为后续入侵实施阶段突破网络脆弱点做铺垫。Mazurczyk等[20]扩大了信息收集阶段涵盖的范围，将传统网络侦察行为划分为4类：互联网情报、网络信息收集、侧信道攻击和社会工程，同时提供具体例子，以技术新旧、交互程度深浅4个维度，将主流的信息收集手段通过图表的形式进行总结，如图9所示。Roy等[21]在Mazurczyk等[20]的分类基础上，进行更深度而细致的总结，考虑Mazurczyk等[20]遗漏的侦察方法，并从更多的角度，提出多种不同的分类方法。

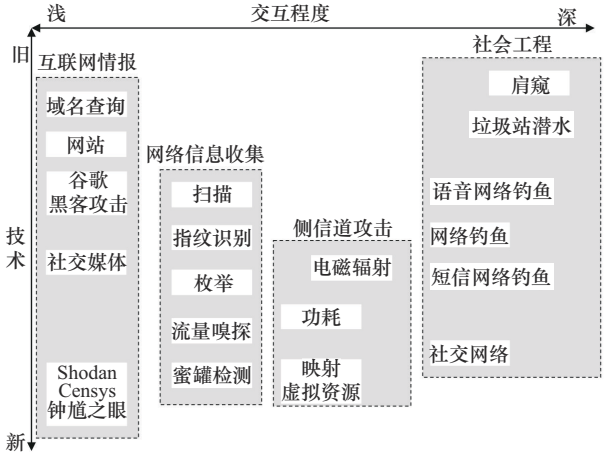


图9　主流的信息收集手段

信息收集技术也随着时间不断地发展迭代。起初，Shaikh等[22]在互联网威胁暴露初期就敏感地指出，随着个人、企业和政府组织对互联网的依赖加深，通过简单地点击鼠标即可发现承载这些敏感信息的服务器。他们通过总结各类主机扫描、端口扫描和脆弱性评估，对早期网络信息收集技术做了总结。Gont等[23]将网络信息收集技术推广到IPv6，进一步拓宽了应用范围。Bou-Harb等[24]重点关注网络扫描技术，全面地总结了前人未提及的目标导向扫描技术和相关隐蔽扫描技术。Salahdine等[25]

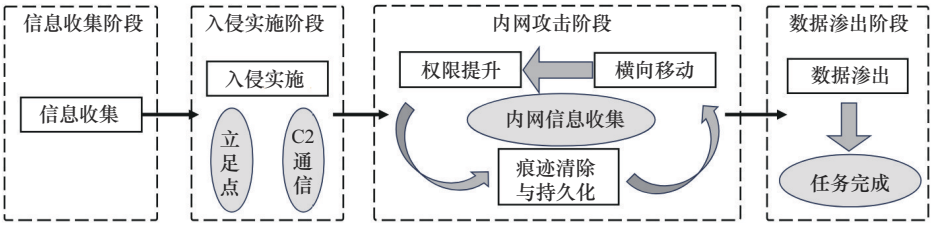


图8　APT全生命周期

系统性地总结了社会工程学攻击，将社会工程学分为针对人和针对计算机的攻击，其中针对人的攻击部分即信息收集阶段的社会工程学技术。

与此同时，基于特定目标的互联网开源情报的信息收集方法也被提出。Kanta[26]通过互联网开源情报收集，应用于更高效的弱口令爆破；Rai等[27]则通过开源的情报技术，寻找特定的Web资产。这一想法也与一些网络测绘工具不谋而合，如Shodan、Zoomeye、Fofa、hunter等。

随着人工智能技术的发展，神经网络辅助的网络侦察技术也逐渐被广大研究人员关注，Millar等[28]第一次将基于图的机器学习方法，引入网络侦察活动，并在真实TCP/IP网络中检验该方法的网络侦察效果，能够得到更深层次的网络关键信息。

信息收集阶段一般不进行实质性的攻击行为，时间跨度较长，是APT活动的重要基础。详尽、全面、广泛的信息侦察能够大大减轻APT攻击的难度，同时在攻击实施过程中也能及时进行信息关联，深化攻击的危害程度。

1.4.2　入侵实施阶段

攻击者在收集了足够的信息后，将利用目标信息系统的薄弱点进行攻击和渗透。该部分的核心是通过对目标信息系统服务器植入恶意代码，使得目标信息系统与C2建立通信，实现立足点的建立。此阶段针对的攻击对象从宏观上可分为目标信息系统的客体（目标组织人、网站管理员、雇员等）和目标信息系统的主体（网络资产）。根据Ussath等[29]的统计，只有少数APT攻击是通过0day进行和实现的，大多数APT攻击都是基于已知的漏洞。

对于客体，APT攻击者主要依靠社会工程学手段，以鱼叉攻击和水坑攻击为主，如Salahdine等[25]的分类，针对计算机的社会工程学攻击即为入侵实施阶段所用到的技术。Center等[17]介绍了一次完整的鱼叉式钓鱼攻击：APT攻击者使用伪造的Mandiant的CEO姓名的邮件，向Mandiant的一名员工发送了一封鱼叉式网络钓鱼邮件，邮件中包含一个恶意压缩文件，目的是安装一个可执行的后门程序“WEBC2-TABLE”，以在目标公司服务器上建立立足点，完成C2通信。

对于主体，开放式Web应用程序安全项目（OWASP）每若干年会发布网络安全风险Top 10，现已成为网络安全最权威的总结，最新版本于2021年发布，其总结的十大风险也是APT攻击者针对信息系统主体经常利用的手段。

值得注意的是，APT攻击者的攻击面不仅仅局限于网络资产，移动设备、工业控制设备以及物联网（IoT, Internet of things）均为APT攻击者的打击范围，OWASP组织也列举了移动设备的Top 10风险、IoT设备Top 10风险、API安全的Top 10风险。随着近年来，大语言模型（LLM, large language model）的兴起，LLM安全的风险也逐渐走入APT攻击者的视野，成为新阶段APT攻击的一大特征。

在入侵实施的过程中，C2通信的隐蔽性往往是APT攻击者关注的重点，这也是APT攻击能够持续很长时间的基础。FireEye公司[30]披露的APT29组织的一种高隐蔽通信手段HAMMERTOSS攻击将恶意代码利用流行合法的Web服务如推特，GitHub传递消息，以掩盖攻击行为。攻击者通过注册指定账户发布推文，受害者PC端的木马访问该账户的推文获得目标统一资源定位符（URL, uniform resource locator）和解密密钥，再从目标URL获得图片结束符后添加的加密文本，利用解密密钥解密，得到对应的恶意控制指令并执行。HAMMERTOSS攻击案例的流程如图10所示。

域前置技术是另一种APT攻击者常用的隐蔽通信手段，攻击者在不同的通信层使用不同的域名，在HTTPS加密下，审查员无法区分域的前置和非前置流量，用这种方法伪装成被允许的知名云平台的流量来通过网络边界审查[31]。

尽管大多数APT攻击活动并未使用0day漏洞，但基于0day的APT攻击并未被厂商感知，暂无漏洞补丁，同时也并未存在漏洞指纹，导致其具有无法检测、危害性大的特点，因此基于0day的攻击仍是不容忽视的一点。国家级的APT组织通常拥有0day武器库，也具备0day漏洞挖掘能力，在入侵实施阶段可能精心构造针对目标信息系统的多组0day攻击链，实施高隐蔽性、高威胁性的攻击行为。

入侵实施阶段将进行实质性的攻击行为，其本质是针对信息系统的初次攻击，与受害主机建立立足点，实现稳定、隐蔽的C2通信，特征较为明显，为内网攻击阶段做准备。

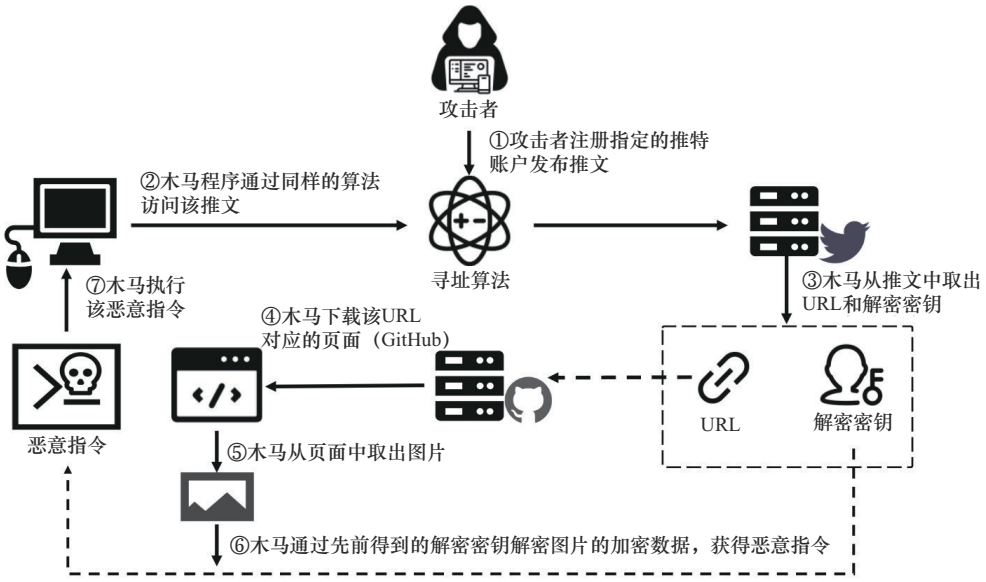


图10　HAMMERTOSS攻击案例的流程

1.4.3　内网攻击阶段

在目标服务器与C2建立通信、完成立足点建立后，APT攻击者便可通过C2服务器向受害主机发送控制命令，此时进入内网攻击阶段。攻击者使用各种技术从受损的系统访问其他主机，并获得敏感资源的访问权[8]。该阶段主要包括3个部分：权限提升、横向移动和痕迹清除与持久化。

1) 权限提升

APT攻击者出于隐蔽性和满足长期控制的需求，需建立高权限，持久化的后门，因此需要进行权限提升操作。权限提升方法具体可总结为：①系统漏洞权限提升；②数据库权限提升；③系统配置错误权限提升；④权限继承类权限提升；⑤第三方软件、库、程序权限提升；⑥WebServer漏洞权限提升。持久化手段具体可总结为：①修改注册表；②开机启动项；③计划任务；④dll劫持；⑤引入第三方软件。

2) 横向移动

出于安全性的考虑，很多企业及组织的关键资源并未对互联网开放，而是仅仅对内网用户开放访问权限，因此APT攻击者需要以立足点为跳板机，向其他内网资产进行横向移动。横向移动过程示意如图11所示，对于有网络边界防火墙的目标系统，攻击者可能通过绕过或者隧道技术，突破防火墙，继续完成攻击活动，横向移动的最终目的是突破隔离网段，扩大攻击成果。

3) 痕迹清除与持久化

出于隐蔽性的要求，痕迹清除与持久化是攻击

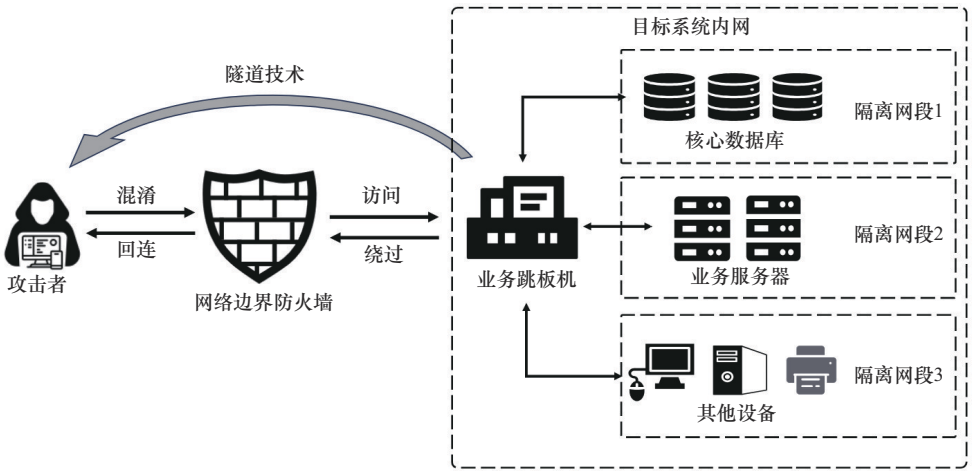


图11　横向移动过程示意

者的必要行为，一方面可以增加溯源反制的难度，另一方面可以避免被受害者感知，延长攻击周期。痕迹清除与持久化环节首先应重点关注各类日志，删除各类日志中记录的关于本次攻击的敏感操作；其次是后门隐藏的持久化方法，如Rootkit等技术隐匿，避免植入恶意载荷的程序被发现。

权限提升攻击的相关研究起源很早，Chari等[32]在早期的文章中关注基于文件名的权限提升攻击，通过创建文件系统链接“诱骗”受害者打开非预期文件。Davi等[33]系统地总结了Android系统的权限提升攻击方法。Suciu等[34]研究可信执行环境（TEE, trusted execution environment）中的水平权限提升漏洞（HPE, horizontal privilege escalation），在真实的三大基于ARM TrustZone可信操作系统上发现了19个水平权限提升漏洞。Kim等[35]重点研究基于浏览器插件的权限提升攻击。可以看到，权限提升攻击的应用范围和概念在不断地拓宽和增加。

实际APT攻击中常用的权限提升通常需要具有普适性和极大的危害，例如常被使用的脏牛（Dirty COW, dirty copy on write）权限提升漏洞。该漏洞于2016年[36]被发现，被广泛应用于Linux和Android内核权限提升，可使无权限用户获得对运行基于Linux操作系统的设备的完全控制权（如root权限），由于其使用的写入复制技术被广泛应用于各类Linux内核系统中，此漏洞造成的影响和危害更加严重。Saleel等[37]在文章中系统的讨论和分析了Linux中的写入复制问题，并解释该问题的性质、产生原因以及缓解该问题的不同机制。同样，Lin等[38]也关注到了被广泛应用于Linux Kernel的权限提升漏洞，基于脏管道漏洞提出了一种新的内核凭证替换技术，该方法将非特权和特权内核凭证互换，从而为该漏洞提供了类似脏管道漏洞的可利用性，实现了权限提升攻击。

在横向移动方面，Haber等[39]从攻击者的视角出发，梳理横向移动攻击的目标，以资源为导向进行探讨。Demers等[40]通过Kerberos的身份验证攻击媒介Kerberoasting，实现系统内的持久性或横向移动。在实际应用方面，Mimikatz是Windows系统中横向移动、域渗透常用的工具，其提供了各类内网密码转储存操作，同时也支持域环境中的伪造黄金票价和白银票据，用于更深度的横向移动攻击。

在痕迹清除与持久化方面，Garcia等[41]将Rootkit技术扩展到IoT设备中，实现了一种名为HARVEY的PLC rootkit，能够在网络物理电网控制系统中进行物理无感知的隐形攻击。Wampler等[42]提出了ExSpectre，将任意恶意代码编译成看似无害的有效载荷二进制文件，误导CPU的分支预测器，实现高隐蔽性的后门隐藏持久化。

内网攻击阶段的攻击环境处于内网中，攻击手法也与前2个阶段有鲜明的差别，通过内网的攻击，能够最大化地达成攻击目的，放大攻击危害。

1.4.4　数据渗出阶段

数据渗出阶段是APT生命周期的最后阶段，APT攻击者在经历信息收集阶段，入侵实施阶段后建立起C2通信，并通过横向移动在目标系统中寻找到目标数据，将收集到的数据泄露到自己的命令和控制服务器。Alshamrani等[8]额外提出，由于大多数入侵检测和防御系统只进行入口过滤而不进行出口过滤，因此实际的数据泄露行为难以被发现。

Ullah等[43]的工作从数据泄露攻击向量的方面，全面总结了各类数据泄露攻击的方式。数据泄露攻击向量如图12所示。

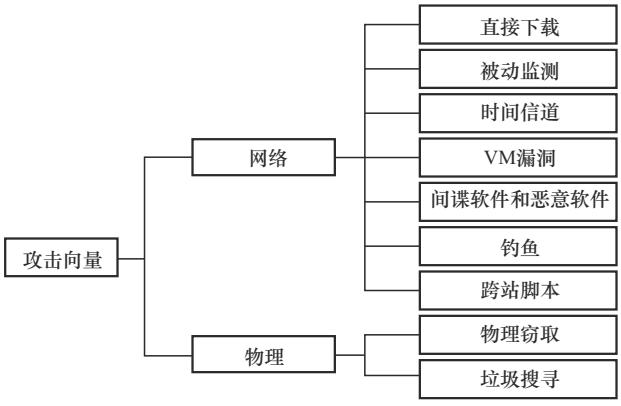


图12　数据泄露攻击向量

Nar等[44]指出数据泄露通常以很缓慢的速率进行，除非攻击者能够一次发送全部数据，并且留在目标系统中没有任何好处。为了避免检测系统的检测，在被窃取的文件发送之前，可以将文件重新格式化，加密或者附加到其他文件之中。例如，在Duqu的案例中以JPEG文件的形式泄露[45]。

数据泄露直接攻击方式大多基于互联网暴露面

的未授权访问以及结构化查询语言（SQL, structured query language）注入。Ullah 等[43]将直接方式仅仅细分为 SQL 注入的攻击方式，并不能很好地总结。基于互联网暴露面的未授权访问包括：未授权的应用程序编程接口（API, application programming interface）泄露、API 水平越权遍历、网站敏感信息备份文件泄露等。值得注意的是，由于调用 API 进行的信息交互本身被认为是合理的请求，且存在高频、高并发的特点，因此该方式导致的信息泄露天然具有一定的隐蔽性。基于 SQL 注入的攻击方式在相关研究[46-48]中已经介绍得十分完备，SQL 注入作为 APT 攻击者最常用的攻击手段之一，与其相关的高集成、高自动化工具如 Sqlmap 提供了非常便利的攻击应用[49]。

间接的信息泄露攻击常用方式如监测与嗅探、钓鱼等，并不直接作用于信息系统主体，而是通过间接的方式，从信息交互的参与方入手进行信息的窃取。Anu 等[50]指出攻击者会在不同的级别通过不同的攻击来窃取不同级别的数据。间接方式的信息泄露攻击广泛地存在于真实世界，Dragos 等[51]从各类网络协议的角度出发，总结了嗅探攻击在计算机网络中的方式与危害。Orazio 等[52]通过研究 IoT 设备中的 iOS 配队模式引发的新型隐蔽式数据泄露攻击，为数据泄露攻击提供了新的攻击维度，同时也将攻击覆盖面扩展到物联网设备领域。Sarkar 等[53]开发了一套针对 IoT 设备中低功率蓝牙的嗅探方案，并在真实开源平台上实现了更高的嗅探精度。

数据渗出阶段是 APT 攻击活动的最后一个阶段，是将敏感数据外带，进行成果确认的过程，该过程存在显著的特点，即信息从高密度向低密度流动，因此，许多数据泄露防护（DLP, data loss prevention）进行核心数据全覆盖，可以有效地防止数据渗出的发生，具体内容将在数据渗出阶段的检测进行介绍。

2 基于 APT 活动全生命周期的检测技术

2.1 信息收集阶段的检测

在信息收集阶段，攻击者为入侵实施阶段建立稳定立足点做准备，会对目标系统进行大量的信息收集。在这一阶段通常采用漏洞或端口扫描工具、网络嗅探等技术获取目标组织的网络资产信息；采用社会工程技术和开源情报等手段获取目标组织的工作人员、运作方式等信息，如图 13 所示。如果能够在该阶段检测出 APT 攻击，就能将损失最小化。

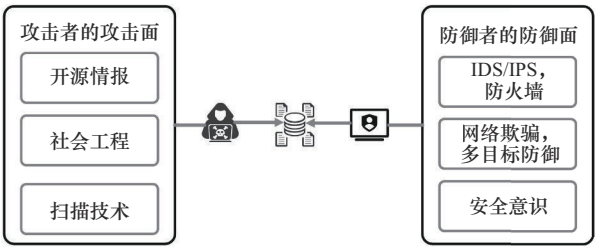


图 13 信息收集阶段的攻击与防御面

1) 主动交互侦察攻击的检测

在信息收集阶段，攻击者可以从远程发起对目标网络的主动侦察，通过扫描技术与目标系统进行交互，分析返回数据来获取相关信息。Panjwani 等[54]通过实验调查指出端口扫描等扫描技术往往是网络攻击的前兆。因此，有效地检测各种扫描行为可以很好地防范网络攻击。

面对主动扫描侦察攻击，防御者可以通过检测识别网络流量来判断网络侦察攻击。Bou-Harb 等[24]指出经典的防御和检测技术：防火墙和入侵检测系统（IDS, intrusion detection system），仍然是抵御扫描侦察的首要方法。Anderson 等[55]第一次详细阐述了 IDS 的概念。Denning[56]则提出了第一个 IDS。早期的 IDS 如 Snort[57]，采用基于签名的检测方法，通过将攻击模式提取为特征并存储在数据库中进行比对。这种方法的优点是易于创建和更新特征，操作简单，适用于阻止特定攻击。但缺点是需要提前了解攻击模式，无法应对新模式的攻击，只能检测已知的扫描攻击。

基于阈值的检测方法也是一种传统的检测方法，它通过预定义的阈值来识别扫描入侵。Heberlein 等[58]设计的网络安全监视器（NSM, network security monitor）被认为是实现基于阈值的扫描检测方法的先驱，主要包括数据捕获、数据分析和支持 3 部分。如果一个源在一段时间内联系超过 15 个 IP 地址，或试图联系一个无响应的 IP 地址，NSM 会将其识别为异常，可能是恶意的源。

基于规则的方法利用知识库中存储的规则来检测和预防典型扫描，例如，基于模糊规则的慢端口扫描检测方法[59]。此外，利用可视化工具来检测

特定网络事件，例如可视化Nmap、Nessus等的扫描模式[60]。强调灵活信息处理以追踪扫描源的软计算方法，例如Chen等[61]提出的一种基于单亲遗传算法的快速端口扫描方法。

随着人工智能的发展，研究人员开始采用人工智能技术来识别扫描行为的模式。例如通过训练模型区分正常网络流量和扫描行为。Almomani等[62]对比了几种增强机器学习分类器在识别侦察攻击中的效果，在UNSW-NB15数据集上的实验表明，Cat Boosting分类器效果优于其他分类器。

2) 基于网络欺骗的检测防御措施

近年来，防御者开始采用网络欺骗（如蜜罐策略）和移动目标防御等措施应对侦察攻击。Crouse等[63]使用概率模型讨论了这些防御措施的重要性，指出移动目标防御通过改变攻击面，使攻击者不能对网络状态做出静态和长期的假设，从而影响侦察攻击。Kimberly等[64]的实验分析表明，相比于没有使用欺骗的情况，诱饵的存在和提供存在欺骗的信息结合对攻击的影响最大。

蜜罐欺骗防御检测方法中，防御者使用诱饵数据或蜜罐网络环境欺骗APT攻击者，并监控蜜罐访问以检测APT攻击。Zhong等[65]提出了一种基于蜜罐技术的方法来防御检测APT信息收集阶的攻击者。他们构建了一个蜜罐防御检测机制如图14所示，其中秘书负责前端交互和传递数据库内容，攻击检测子系统利用监督式机器学习算法识别攻击者，并向调度员传递预警信息。调度员决定向秘书交付真实数据还是带有蜂蜜印记的欺骗数据，后者是通过生成对抗网络（GAN, generative adversarial network）生成的，可以避免攻击者获取关键信息。当交付欺骗数据时，数据库保留相应记录，以监控对欺骗数据的访问，进一步跟踪检测APT攻击者。

3) 被动方法侦察攻击的防御

当攻击者通过第三方进行侦察，例如使用开源情报被动方法手段，防御者将无法检测到。

Yu等[66]对医院网站的隐私和安全问题进行了详细的调查研究，指出绝大部分的医疗机构网站存在隐私信息泄露问题。Rugo等[67]探讨了公共机构公布的信息可能被APT攻击者利用的问题。他们模拟了对意大利某机构的攻击，利用其公布的信息获得了完整的技术与供应链清单，并利用清单中信息成功绕开了机构的防御措施。

因此，防御者需加强隐私信息发布的安全意识，通过制定严密的信息发布流程和设立公共信息发布监控平台等措施提高防御水平。

4) 针对目标人员侦察攻击的检测与防御

信息收集阶段的另一大突破口是目标系统的工作人员，APT攻击者可以使用钓鱼、欺骗和威胁等手段利用内部工作人员获取关键信息。

APT攻击者可能会伪装成客户通过聊天软件诱导工作人员泄露关键信息。Tsinganos等[68]提出了一种利用卷积神经网络和自然语言处理（NLP, natural language processing）技术识别基于聊天诱导的社会工程攻击的方法。他们在一个基于语言诱导原则标注的社会工程语料库上训练了一个网络，使其能在聊天中检测出诱导泄露信息的意图，从而保护工作人员。

随着社交媒体的发展，人们在社交媒体上泄露了大量可能危害隐私的信息。为了解决这个问题，许多研究者开始检测和分析社交媒体和博客上的自我表露行为。Peiretti等[69]建立了一个意大利语网

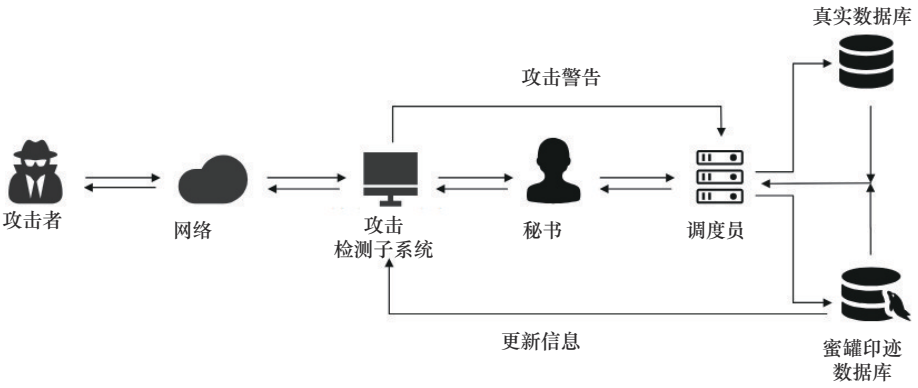


图14　蜜罐防御检测机制

络帖子语料库并应用几个基于Transformer的语言模型，根据它们的敏感度对它们进行分类，在测试中表现出了良好的分类效果。

网站式钓鱼攻击则通过设立恶意网站，等待工作人员的访问，窃取他们的信息。Rajeswary等[70]指出，恶意网站的检测方法可以分为基于URL、内容、声誉和行为的检测，这些技术可以组合使用以提高准确性。Kim等[71]基于大多数攻击者会隐藏痕迹和成本的原因，重复使用同一个钓鱼网站，选择不需要个人信息的网页托管等行为特征，提出了一种基于网络推断的方法，能精确检测伪装成合法模式的钓鱼URL，对其规避行为具有很强的效果。

值得一提的是，培养工作人员防范社会工程攻击的意识也是重要的措施。Hafner等[72]提出了一款社会工程攻击的桌游，通过这款游戏让参与者得到训练，提高对社会工程攻击的防范意识。

信息收集阶段检测方法对比如表1所示。

表1　　信息收集阶段检测方法对比

分类	方法	特点
主动交互侦察攻击检测	文献[54]	扫描流量
	文献[55-57]	IDS
	文献[58]	基于阈值
	文献[59-61]	基于规则
	文献[62]	基于AI
基于网络欺骗的检测	文献[63]	概率模型
	文献[64-65]	蜜罐检测
被动方法侦察攻击防御	文献[66-67]	信息泄露
针对人员侦察攻击检测	文献[68]	聊天信息
	文献[69]	媒体信息
	文献[70-71]	钓鱼网站
	文献[72]	人员意识

2.2　入侵实施阶段的检测

入侵实施阶段中，APT攻击者利用信息收集阶段收集到的有利于系统入侵的关键信息，对目标系统展开入侵。

1) 鱼叉式钓鱼攻击、水坑攻击的检测防御

鱼叉式钓鱼是当今互联网中一种突出的有针对性的攻击方式。攻击者通过前期收集的信息，精心伪造定制的邮件，冒充可信的电子邮件发件人，诱骗受害者启动包含恶意代码的附件或点击恶意链接，从而使攻击者在保护良好的网络中获得立足之处。Duman等[73]提出通过对邮件元数据和邮件内容的文体特征建立概率模型，检测鱼叉式钓鱼攻击的特征指标，从而区分鱼叉式钓鱼攻击和合法电子邮件。随着人工智能技术的发展，现阶段的检测手段更多采用机器学习和深度学习方法对钓鱼邮件和真实邮件进行分类检测。

Cidon等[74]针对商业电子邮件系统，提出了一种用于检测商业电子邮件攻击的检测器BEC-Gaurd。Bai等[75]关注基于社会工程学的电子邮件钓鱼攻击检测，他们总结了电子邮件业务中APT攻击的战术、技术和程序，并利用深度学习算法学习深层特征进行APT电子邮件检测。

尽管防御者对鱼叉式钓鱼攻击进行了充分防范，但攻击者也发展了技术，通过损害企业白名单中的第三方来渗透目标企业网络。攻击者通过前期信息收集，获取目标组织成员经常访问的网站，选择易植入恶意程序的网站，等待目标组织人员访问，实现对目标组织网络的初步入侵，这种攻击被称为水坑攻击。水坑攻击流程如图15所示。

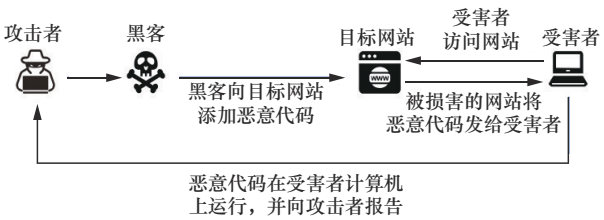


图15　水坑攻击流程

Allen等[76]指出，关于检测或调查复杂水坑攻击的研究相对较少，大多数研究旨在检测对良性网站的恶意修改。他们提出了Mnemosyne，一种基于浏览器攻击起源的事后取证分析引擎，可以准确地重构、调查和评估水坑攻击的后果。Mnemosyne通过用户级分析来评估恶意修改对目标企业的影响，并识别受害者，从而帮助企业更好地防范水坑攻击并减少损失。

2) 利用漏洞的入侵检测与防御

APT攻击者可能会利用已知漏洞入侵目标组织网络系统。攻击者通过信息收集，获取目标组织的软件和操作系统版本信息，然后在通用漏洞披露（CVE, common vulnerabilities and exposures）和国家漏洞数据库（NVD, national vulnerability database）等漏洞数据库中查找相应的漏洞，尝试利用这些漏洞进行入侵。因此，防御者需要及时更新安全补丁，以防止此类攻击。

3) 面向恶意软件的检测与防御

当前述攻击手段执行成功后，APT攻击者往往会植入恶意软件，并等待其的运行，打开通往目标系统的大门。攻击者在此时会面临目标系统对于恶意软件检测的挑战。Aslan等[77]将恶意软件的检测分为3个阶段：恶意软件分析、特征提取、基于特征检测。他们将检测方法分为基于签名、行为、启发式、模型检查、云、深度学习、移动设备和物联网等不同类型。并总结了各种方法的优缺点。

4) 面向C2通信的检测

APT攻击者在成功运行恶意软件后，下一步是建立C2通信通道，为了避免被发现，通常使用常用协议伪装成常规流量。Wang等[78]指出相较于其他协议，超文本传输协议（HTTP, hyper text transfer protocol）被广泛用于C2通信，因为基于HTTP的C2流量在大多数企业中被视为合法，而其他协议如点对点（P2P, Peer-to-Peer）协议和网络中继聊天（IRC, internet relay chat）协定流量具有明显的网络特征。这提示我们可以根据不同协议的特点设计C2通信通道检测器。

Talib等[7]将检测C2通信的方法分为4类。①基于行为或网络的检测，其依赖于主机行为或网络流量的异常，如网络时延过大、流量过大、异常端口流量和异常系统行为。该策略旨在检测与正常网络活动不一致或与C2活动相似的行为（例如，Yan等[79]在分析域名系统（DNS, domain name system）请求记录的基础上，提出一种新的特征来表示DNS请求和响应消息之间的关系，并根据计算出的可疑值对检测到的DNS行为进行威胁评估）。②基于签名的检测，利用从C2通信中提取的预定义的模式来检测（例如，Menon[80]通过跟踪单个进程的DNS响应，将具有异常大量域名解析失败的进程列入黑名单）。③基于图的检测，通过在相关异常模式之间引入连接，图自然地描述了它们之间的联系（例如，Tran等[81]提出了一种利用图推理，基于域名与客户端关系来计算域名恶意评分的方法）。④基于深度学习的检测，将网络流量转换为特征向量，将其传递给神经网络模型，从空间和时间维度提取不同的模式，利用卷积神经网络（CNN, convolutional neural network）或长短期记忆（LSTM, long short-term memory）网络等学习网络流量特征。Highnam等[82]创建了一个新的混合神经网络，Bilbo模型，Bilbo模型是第一个并行使用CNN和LSTM网络进行域名生成算法（DGA, domain generation algorithm）网络检测的模型，测试显示在4 h的实际网络流量中，该模型发现了至少5个潜在的C2通信。Yun等[83]进一步提出了一种新型的恶意通信流量检测模型（HMCD-Model, http-based malicious communication traffic detection model），该模型结合了带梯度惩罚的瓦瑟斯坦生成对抗网络算法（WGAN-GP, Wasserstein GAN with gradient penalty）和基于CNN和LSTM网络的混合神经网络，能够有效检测未知的基于HTTP的恶意通信流量。

入侵实施阶段具有明显的流量特征，从传统的检测方法到最新研究中，大多是针对该阶段的特征进行检测，也因此最新研究大多集中于基于深度学习进行的流量特征检测，相关检测方法对比如表2所示。

表2　　入侵实施阶段检测方法对比

分类	方法	特点
钓鱼、水坑攻击检测	文献[73-75]	邮件检测
	文献[76]	恶意修改检测
恶意软件检测	文献[77]	阶段划分
C2通信检测	文献[78]	基于协议
	文献[79]	基于行为
	文献[80]	基于签名
	文献[81]	基于图检测
	文献[82-83]	基于深度学习

2.3　内网攻击阶段的检测

内网攻击阶段是APT攻击者扩大攻击成果的阶段，该阶段会深入受害系统的内网以寻求更多敏感信息及目标，该阶段攻击方法特征较为明显，现有研究针对该阶段的检测也已形成完备的体系，涵盖权限提升、横向移动与痕迹清除与持久化3个方面。

2.3.1　权限提升的检测

Yamauchi等[84]针对操作系统的权限提升漏洞的检测，开发了一种可以防止利用操作系统漏洞的特权升级攻击的内核观察器AKO。具体来说，该观察器监控系统调用前后权限的变化，并将其视为权限提升攻击进行防范。Elsabagh等[85]关注Android系统的固件镜像权限提升漏洞问题，开发了一种FIRMSCOPE框架，用于扫描Android系统权

限提升漏洞。在100多个安卓供应商预装的应用程序中，FIRMSCOPE发现了850个权限提升漏洞，且大多为可利用的0day漏洞，效果异常卓越。Suciu等[34]在研究可信执行环境中的HPE后，设计了一个基于符号执行的HPE漏洞自动扫描程序HOOPER，用以检测该特征的水平权限升级漏洞。Li等[86]将模糊测试技术应用于Android系统权限相关漏洞检测中，设计了一种黑盒模糊测试工具CuPerFuzzer，用于检测Android系统中的悬挂自定义权限、权限组映射不一致、自定义权限提升和权限定义不一致等缺陷。

2.3.2 横向移动的检测

Tian等[87]提出了一种名为CloudSEC的实时横向移动检测方法，基于边缘云环境证据推理网络，以防止在边缘云计算环境中进行横向移动攻击。该方法引入了漏洞相关性的概念，构建证据推理网络，利用网络系统的漏洞知识和环境信息来检测横向移动攻击。Yan等[88]通过建立理论框架描述了APT对内部网络的攻击，选择切入点与切入模型研究整个APT周期的行为，进行模拟观察目标攻击模型的动态变化，以验证APT攻击的特征。Mohamed等[89]构建了一种新模型，基于对抗性战术技术和尝试（ATT & CK, adversarial tactics, techniques, and common knowledge）矩阵，用于检测第一个潜在受害者遭受的APT攻击。该奇怪行为检测（SBI, situation behavior impact）模型调查和监控CPU、RAM、Windows注册表和文件系统中的APT行为特征，也可用于横向攻击检测。

一些学者通过提取APT攻击的特征，针对特定受害群体，研究横向移动检测手段。

Bai等[90]专注于Windows系统的横向移动攻击检测，通过远程桌面协议（RDP, remote desktop protocol）调用为特征，开发了一种异常检测方法用于检测横向移动，并与多种有监督的机器学习技术评估了该方法的优越性。除了网络资产的横向攻击，工业物联网因其战略价值、内网资产众多，成为APT攻击的重点目标。Bi等[91]着重研究工业物联网内网的横向移动攻击，建立了节点级的状态演化模型，计算了每个设备被APT攻击入侵的概率。

2.3.3 痕迹清除与持久化的检测

具体分为日志系统破坏检测和隐藏后门的检测。Paccagnella等[92]在2020年提出了基于内核的日志防篡改系统——KennyLoggings，确保同步的完整性，并保留日志防篡改的证据。同年，Paccagnella等[93]再次提出了基于可信执行环境的日志防篡改系统CUSTOS，提高了防篡改日志的记录效率，并保持了安全性。Jiang等[94]针对APT攻击者可能破坏系统审计框架的问题，提出了日志审计框架NODROP，通过隔离不同进程生成的系统调用日志，确保日志完整性，有效避免了攻击者对审计日志系统的清除行为。Sekar等[95]关注持久化日志储存中的日志篡改问题，通过伯克利数据包过滤器（eBPF, extended berkeley packet filter）框架进行优化调整技术，减少了日志数据的工作负载强度，并缩小了日志篡改数据面。

在隐藏后门检测方面，Tian等[96]提出了基于硬件辅助虚拟化技术的内核rootkit检测系统VKRD，为目标内核模块提供透明、高效的执行环境，更好地检测恶意内核模块。Nagy等[97]通过可信执行环境（TEE, trusted execution environment）进行IoT设备的Rootkit检测，提出了Rootkit检测算法和持久储存中的Rootkit组件检测算法。Yao等[98]关注Web中的恶意软件应用，开发了自动恶意软件分析工具Marsea，通过分析恶意软件样本，进行网络应用程序恶意软件检测。

本节总结的内网攻击阶段的检测方法具体如表3所示。

表3 内网攻击阶段检测方法总结

检测阶段	方法	应用方面
权限提升	文献[84]	Linux
	文献[85-86]	Android
横向移动	文献[87]	云计算环境
	文献[88]	内部网络
	文献[89-90]	Windows
	文献[91]	工业物联网
痕迹清除与持久化	文献[92-95]	日志防篡改
	文献[96-98]	后门检测

2.4 数据渗出阶段的检测

数据渗出阶段，是APT攻击活动的最后一个阶段，也可以视为针对APT攻击的最后一道防线。攻击者会尽可能保持低调，把大规模的数据伪装成正常流量通过C2通道离开网络。

有很多手段可以预防数据渗出，比如最为常用的文本分类、实施访问控制[99]，以及身份验证机

制限制对敏感数据的访问，使用加密技术对敏感数据进行加密等。预防措施并不能完全阻止数据渗出，本文重点关注的是数据渗出的检测手段，需要推断数据传输的企图，判断出可能的数据渗出并阻止他们。

数据渗出阶段检测的难点主要体现在以下几方面：加密数据的检测、准确性问题、误报与漏报、数据流量大和新型攻击技术检测困难。

随着黑客攻击技术的不断进化，他们可能采用新的数据外泄手段和技术，这对数据外泄检测带来了额外的挑战。

现有的数据渗出检测技术，主要分为基于流量内容的检测方法与基于流量行为的检测方法2类。

2.4.1　基于流量内容的检测方法

内容检测方面，通过检测传输数据的内容，分析其关键字等信息，防止未经授权的数据泄露。

DLP系统试图通过监控信息传输的内容或上下文来防止有意或无意泄露敏感信息[100]。Alkilani等[101]概述了在多种场景中使用的基本数据泄露技术，并且使用多种机制来检测数据外渗，每种检测机制都能提供独特的功能，用于在消息或文档中检测违规行为。Kiperberg等[100]搭建了一种基于虚拟化技术的高效数据泄露防护系统，用于防止内部人员通过恶意行为或社交工程等方式泄露敏感数据，并介绍了内容匹配和上下文匹配2种数据泄露防护系统的工作原理和特点，以此来判断是否存在数据泄露行为。

攻击者为了避免检测，会隐藏数据以绕过数据丢失防护系统。隐写技术是一种重要的方法，可将任何敏感内容隐藏在如图像、视频和文本文件等普通数据中。Almomani等[102]使用混合加密隐写术方法，把信息隐藏在高分辨率视频帧中。

目前已经提出多种方法来检查被隐藏的内容，以检测数据泄露的企图。Fathi-Kazerooni等[103]提出了一种新颖的基于CNN的检测方法，依靠手工制作的特征空间作为CNN输入层，搭建了一个隐写分析检测器，提供高达94.3%正确分类测试数据样本。Płachta等[104]分析了在图像隐写检测中采用各种浅层和深度学习算法的性能，证明，集成分类器是基于深度学习的检测方法的一个较好方案。Zhai等[105]针对视频隐写技术，提出了一个12维的通用特征集，并进行了大量的实验，证明其能够检测多个领域的视频隐写术。Hu等[106]对语音隐写技术提出了一个称为隐写分析特征融合网络（SFFN, steganalysis feature fusion network）的深度模型，SFFN可以有效地提取异构并行隐写术（HPS, heterogeneous parallel steganography）中使用的隐写方法的隐写分析特征，并可以融合这些特征以做出可信的预测。

2.4.2　基于流量行为的检测方法

网络流量异常检测是指通过监控网络流量，识别出与正常流量行为不符的流量，从而发现网络中的异常行为。可以通过网络流量异常检测及时发现数据渗出行为，阻止APT攻击，也可以进一步追踪数据流向。

目前，有多种方法可以用于网络流量异常检测，例如，基于规则的方法、基于统计的方法和基于机器学习的方法。其中，基于机器学习的方法是目前最为先进和有效的网络流量异常检测方法之一。

Nadler等[107]提出了一种检测DNS上的隧道和低吞吐量数据泄露的方法。收集每个域的DNS日志，并根据每个域的查询行为提取特征，使用异常检测模型对被分类为用于数据泄露的域的DNS请求进行拒绝。Crespo等[108]通过对轻量级协议流量数据进行分析，成功使用机器学习的方法检测SQL注入攻击。Nyakomitta等[109]提出一种使用信息熵分离明文和加密流量的算法，通过启发式扫描观察网络的流量行为，最后使用基于决策树的相关性函数拉结并检查流量。Wen等[110]针对政务云服务API的安全性问题，使用循环神经网络进行API调用分析，以检测非法利用和泄露。郭嘉琰等[111]提出一种基于图神经网络的异常检测算法，引入图结构、属性和动态变化的信息，以学习进行异常检测的表示向量。该算法显著优于传统的网络表示学习算法，提高了异常检测的准确度，并挖掘了网络中存在的有实际意义的异常。Chung等[112]使用交互式机器学习的方法训练模型，在领域专家和机器学习算法之间建立有效的协作，解决领域专家能够识别恶意用户但无法处理大量数据的问题，大大减少了误报和漏报。Marques等[113]提出了一种新颖的多代理数据泄露检测器架构（MADEX），模仿人类免疫系统中存在的机制和特征，使用代理收集的跨网络流量信息，来有效识别由恶意软件执行的数

据泄露活动。Dijk[114]比较了几种人工智能模型，如堆叠式自动编码器、循环神经网络和一类状态向量机。这些模型显示了在数据泄露阶段检测方面的显著提升。最后，介绍了通过分析网络流量的有效负载来成功检测数据泄露的方法，以改进数据泄露检测。

数据渗出阶段因其隐蔽的手段和层出不穷的技术导致检测的困难，网络的规模及其所使用的设备数量也增添了检测的难度，随着人工智能技术的发展，数据渗出检测面临困难将被解决，该阶段相关检测方法对比如表4所示。

表4 数据渗出阶段检测方法对比

检测阶段	方法	特点
基于流量内容	文献[100-101]	DLP
	文献[103-104]	图像隐写检测
	文献[105]	视频隐写检测
	文献[106]	语音隐写检测
基于流量行为	文献[107]	基于DNS
	文献[108]	SQL泄露检测
	文献[109]	启发式扫描
	文献[110]	API泄露检测
	文献[111-113]	神经网络方法
	文献[114]	代理流量检测

2.5 全生命周期阶段检测

随着APT攻击检测技术的发展，单一APT活动生命周期阶段的检测可靠性面临质疑，相关研究更多的是将整个生命周期纳入检测范围进行检测，以确保检测的真实性和可靠性。

2.5.1 基于溯源图的APT攻击检测

Hossain等[115]提出的SLEUTH方法，首次利用因果关系跟踪和溯源图构造模型以重构APT攻击，与平台无关，基于主存的审计日志数据依赖图抽象，可以在企业主机上实时重构攻击场景，实现实时检测，具有较高的运行效率与检测精确度。

Milajerdi等[116]提出Poirot方法，利用网络威胁情报（CTI, cyber threat intelligence）关联性构建查询图以检测APT攻击，利用内核审计日志构建溯源图，将威胁检测建模为非精确的图模式匹配（GPM, graph pattern matching）问题，创新性地引入图匹配算法，能有效从溯源图中实现APT组织查询图（攻击链）匹配及对齐。该团队于同一年又提出了HOLMES方法[117]，有效利用攻击活动可疑信息流的相关性，将APT活动信息映射到杀伤链，设计高级场景图（HSG, high-level scenario graph）解决低层次信息（溯源图）映射到高层次（攻击链）的语义鸿沟问题，引入降噪算法解决HSG紧密性问题，构建出一种实时检测APT攻击的系统，能有效检测长期潜伏实时的APT攻击。

Han等[118]设计了一种基于溯源图的运行时APT检测方法UNICORN，能在没有先验攻击知识的前提下实现APT攻击检测，准确率高且误报率低。这是第一个对本地完整系统进行运行分析的APT入侵检测系统，且引入了直方图与概要图来对抗长时间潜伏的投毒攻击。

Zhao等[119]提出一种基于异构图卷积网络的威胁情报模型HINTI，用以建模失陷指标（IOC, indicator of compromise）之间的依赖关系，从非结构化威胁描述中自动提取网络威胁对象，并使用攻击偏好建模方法，聚集相同偏好的攻击。与现有的CTI框架不同，HINTI旨在实现一个CTI计算框架，以有效提取IOC，并建模和量化它们之间的关系。

Satvat等[120]提出一个自动化工具EXTRACTOR，利用自然语言处理自动从CTI报告中精确地提取攻击行为信息，并使用语义角色标注进行语义分析以理解攻击行为关系，将非结构化文本转化为溯源图，并捕获系统级因果关系。但该方法由于NLP复杂性，在提取精度上有所损失，且难以识别某些未知实体。

Satvat等[121]提出ATLAS方法，利用审计日志生成端到端攻击故事，使用因果关系图、自然语言处理构建基于序列的模型，可恢复攻击关键步骤和攻击故事。此研究还发现，攻击使用的抽象攻击策略所利用的漏洞和执行的有效载荷无关。

Anjum等[122]提出ANUBIS机器学习技术，它使用事件轨迹（由父子关系相关的事件序列）生成的溯源图数据进行训练，贝叶斯神经网络进行分类，根据模型提供的不确定性分数来解释预测结果，若分数较低，ANUBIS会匹配训练集中最相似的示例，以解释预测结果。结果表明，ANUBIS能高准确性、高精度与低误报率地检测APT攻击。

Mahmoud等[123]提出了APTHunter系统，可在早期阶段检测APT。该系统以内核审计日志为可靠的系统活动来源，将威胁猎杀建模为行为匹配

问题，使用溯源图提供系统实体之间的因果关系和信息流，并将威胁指标及其之间的关系表示为溯源查询。该系统在DARPA的对抗性参与和真实世界的APT活动中进行了评估，取得了良好的结果。

综上所述，基于溯源图的APT检测研究大多集中于在溯源图构建的关系网基础上，结合网络威胁情报、卷积网络、自然语言处理、深度学习与神经网络等技术，逐渐提高检测的精确性与效率与检测能力（如在早期阶段检测APT），且已经取得了一定的成效，是APT检测的重要方法。

本节总结的基于溯源图的APT攻击检测方法如图16所示。

2.5.2　其他方式的APT攻击检测

除了溯源图神经网络带来的APT检测方法，还有很多基于全生命周期的检测方法未能系统性地形成脉络清晰的发展路线，但对于APT检测方法的贡献是不容忽视的。

Jia等[124]建立了基于攻击树的APT检测模型，在集中控制的软件定义网络（SDN, software defined network）中检测潜在的APT攻击。Kumar等[125]同样以攻击树模型辅助检测，模拟了针对工控系统的3种著名的APT攻击，即Stuxnet、Blackenergy和Triton，按照攻击树建模语言以系统化和结构化的方式描述了每种攻击，并给出了这类APT攻击的组合特征。

Yi等[126]提出了一种IDS，利用决策树和梯度提升算法检测APT全生命周期各个阶段的APT活动。此外，该模型还通过优化APT阶段或攻击路径来生成APT指纹，帮助模型进行早期APT检测。该模型使用Dataset APT（DAPT）2020进行评估和验证。所提出的模型证明可以有效地对APT活动进行分类，在大多数APT阶段的准确率超过97.63%。此外，该模型在生成APT指纹方面被证明是有效的。

在CTI的衍生研究中，Gao等[127]基于外部开源网络威胁情报知识（OSCTI, open-source cyber threat intelligence），提出了ThreatRaptor系统，该系统可利用OSCTI进行计算机系统中的威胁搜索。Ji等[128]同样基于开源网络威胁情报，开发了ThreatQA系统，该系统通过构建网络威胁知识库，以自然语言处理的方式来促进网络威胁知识的获取，为企业或机构预防APT攻击提供保障。

实际的工业界中，网络威胁情报共享也已有应用。中国科学院计算机网络信息中心开发的网络安全威胁情报共享平台（CNTD）集成了相关互联网企业共建威胁情报共享的平台，是工业化应用的典范。各大安全企业也有各自的威胁情报中心，如IBM公司的X-Force威胁情报指数、奇安信威胁情报中心、360情报中心等，已广泛地进行商业化落地应用。

2.6　小结

在APT检测方面，本文总结攻击方法在各个生命周期阶段所展现的差异性，结合传统方法与新型的人工智能方法，总结各个阶段的检测方法，以图表形式进行汇总，如图17所示。

3　APT活动的挑战与机遇

随着计算机技术的发展，APT攻击与检测技术在新计算机技术下快速迭代发展。新型APT攻击手段不断地产生，带给所有信息系统巨大的风险与

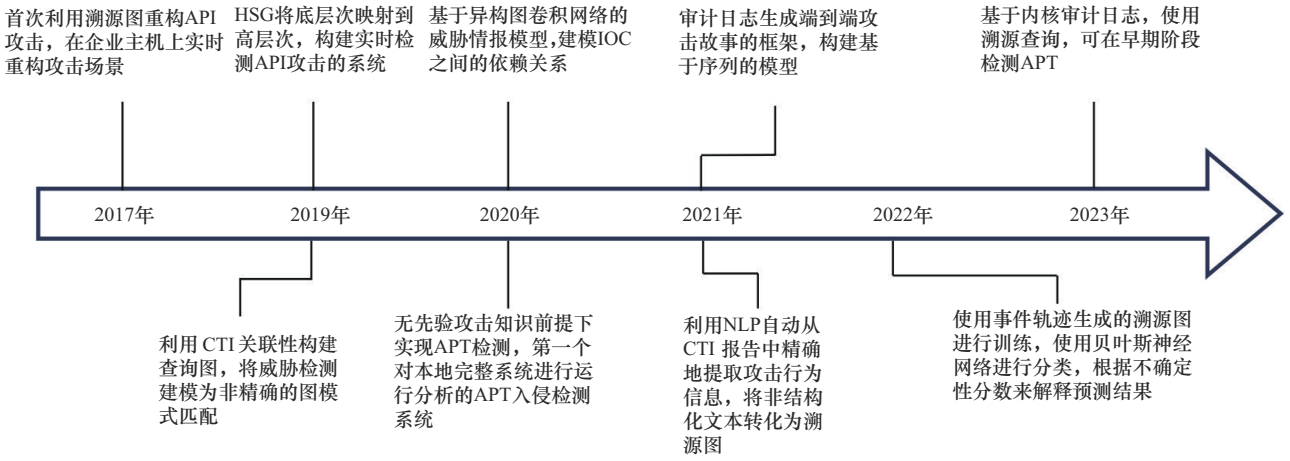


图16　基于溯源图的APT攻击检测方法

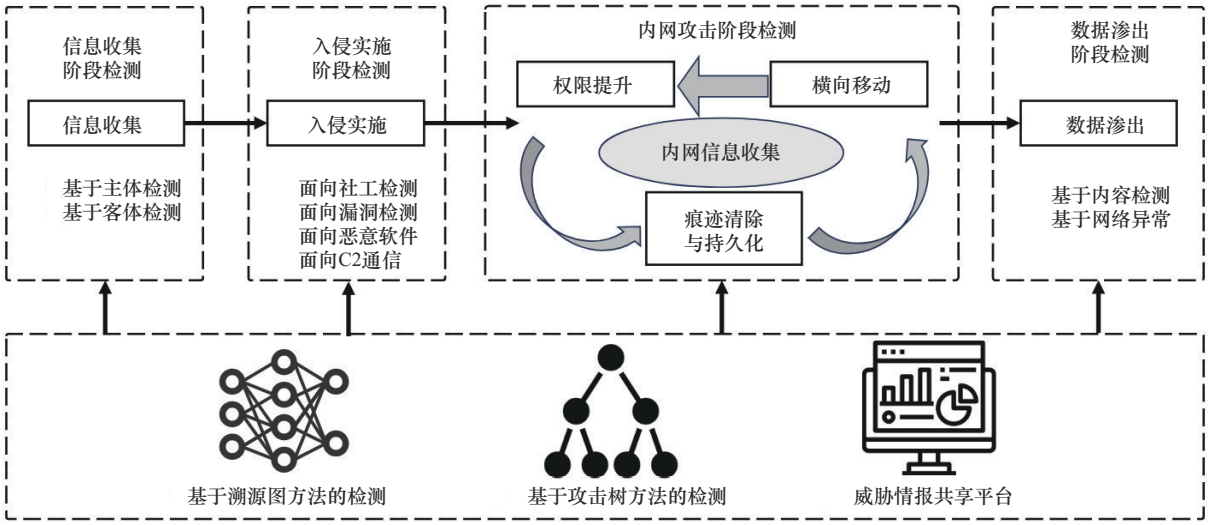


图17　基于全生命周期的检测技术架构

挑战。而随着各类APT检测技术的发展，人工智能的辅助以及威胁情报共享平台的应用，APT检测准确率和鲁棒性快速提高，APT攻击的技术详情透明度不断增加，也带给APT攻击者以挑战。

1) 目前面临的挑战

对于攻击者来说，有以下几个方面的挑战。①闭源的专有系统不断增加，大大增加了攻击者进行源码审计的难度；开源软件的漏洞申报体系不断完善，代码安全重视程度日益增加，促使软件安全性和鲁棒性不断增加，同样增加攻击者的利用难度。②重要信息系统的安全防护非单一化，逐渐产生多维化趋势，从各类入侵检测系统、入侵防御系统、防火墙、数据库审计系统和安全策略加载，敏感系统层层防护；再到威胁情报共享平台标记APT攻击者的攻击指纹与恶意C2服务器，不仅增加了APT攻击者攻击的难度，也缩减了APT攻击的有效生命周期，给持久化与隐藏攻击带来困难。③更安全的加密方式广泛应用，如多方安全计算（SMPC, secure multi-party computation）[129]、全同态加密（FHE, fully homomorphic encryption）[130]等应用于云环境数据交互，敏感信息加密储存，信息交互过程以严格的密文传输，使得嗅探等传统信息泄露攻击手段彻底失效。硬件安全模块（HSM, hardware security module）[131]的方法专用于硬件设备，用于存储和管理加密密钥，提供了更高级别的物理和逻辑安全，以保护密钥免受恶意攻击。

对于受害者来说：①人作为信息系统最薄弱的环节，基于社会工程学的鱼叉式钓鱼攻击和水坑攻击仍然作为最广泛、有效的APT攻击手段，难以严格地从技术层面完全避免。②不安全的API造成的泄露中，攻击者所使用的分布式低速API调用方式进行的数据窃取，难以形成有效的检测手段，几乎无特征与正常API调用进行区分，且使用该方法调用API进行的数据查询很难被安全防护产品感知。③APT攻击者同样使用加密方式进行通信，隐藏在合法通信数据中以规避检测。④无文件攻击技术广泛被APT攻击者所使用，不留下可检测的文件，将恶意代码混淆在正常代码中，使传统的检测方法无效，Lee等[132]总结了相关无文件攻击的攻击技术和特征。新环境下的无文件攻击将会是APT检测研究的另一个挑战。

2) 未来研究机遇

由于APT攻击与检测领域的快速迭代特性，现有的APT的攻击与检测技术与技术发展密切相关。

攻击者应用新型技术，提出更加新型的攻击手法，达到攻击目的；受害者针对新型的攻击方法，进行迭代加固。这将会是常态化的过程，且从一般情况来说，存在着APT检测技术滞后于攻击技术的特点。通过对现有技术的分析，本文也总结了APT攻击和检测中可能会进行的研究方向，具体如图18所示。

从APT攻击的角度分析，有以下几个发展机遇。①0day漏洞组合攻击：攻击者可能会继续寻找和利用新的0day漏洞来入侵目标系统，随着目标系统防御手段的增加，组件供应链的完善，单一的

0day无法顺利地攻破目标系统，往往需要多枚0day漏洞进行组合攻击，构造0day漏洞链。这种攻击方式对攻击者能力要求更高，也对检测带来了更大的困难。②人工智能和机器学习：基于人工智能和机器学习技术来增强攻击能力，现有的研究已经显示，ChatGPT等人工智能应用可以替代人工生成特定的社会工程学钓鱼内容，同时，其本身还会产生如提示词注入等新型攻击方式，随着人工智能的加速应用，不仅仅可以用于辅助APT攻击，针对该类应用的攻击也将成为一大研究方向。③物联网、车联网、云环境等多维度打击面：随着各种专用网络的广泛应用，带来了更多的资产暴露面，攻击者的攻击打击面将不再局限于单一互联网，且信息互联的作用下，根据木桶理论，攻击者可以针对相对薄弱的环节，从而突破目标系统。

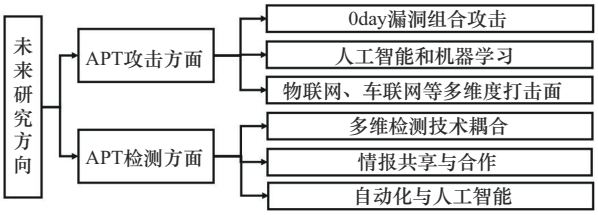


图18　未来研究方向

从检测技术的角度来看，APT检测可能出现以下发展趋势。①多维检测技术耦合：传统的检测方法可能更加适用于某种APT攻击。但是，随着APT攻击的复杂性增加，多维度检测技术的出现迫切需要，基于各种行为、特征、阶段，情报以及漏洞指纹的检测技术将会相互耦合，同步进行APT攻击检测。②情报共享与合作：由于APT攻击往往跨越不同的机构和国家，且攻击者使用的手法和漏洞在一定程度上会有复用的现象，因此，建立情报共享和合作机制将更加重要。基于共享APT攻击的情报信息，可以更早地发现和应对攻击活动。③自动化与机器学习：自动化和机器学习技术的应用将使得APT检测更加智能和高效，快速发展的人工智能技术也将不断促进APT检测，成为一大研究重点。

综上，未来的APT攻击技术势必会趋于更加复杂和智能化，与之对应的检测技术也将趋于完善，两者的协同进化离不开相互之间的研究，对于攻击者，针对检测技术的着重点，混淆流量特征，加密绕过，能够更好地实现攻击目标；对于受害者，第一时间感知到正在发生的APT攻击事件，还原其攻击手法和利用链，有助于实现更加完备优越的检测方法。计算机技术不断迭代的过程中，APT攻击与检测技术也势必更加快速发展，常态化、持续化的对抗研究将是攻击与检测共同的特点。

4　结束语

APT攻击自首次被提出以来，众多研究工作聚焦其攻击特性解读和检测技术研发，在计算机技术的快速迭代下，研究分支冗杂、碎片化，缺乏系统性的归纳与整理。

本文在前人工作基础上，提出更为科学的APT全生命周期划分，并以此为脉络梳理APT攻击与检测技术，最终结合当下研究热点，提出APT攻击与检测技术面临的挑战与研究方向。

APT攻击与检测技术的对抗性发展，本质上也体现着“攻防对抗”的思想，技术的对抗演化过程将是长久的进行时。

参考文献：

[1] KUZNETSOV I, PASHKOV V, BEZVERSHENKO L, et al. Operation triangulation: iOS devices targeted with previously unknown malware [R]. 2023.

[2] Qi An Xin Technology Group Inc. Global advanced persistent threats (APT) mid-2023 report[R]. 2023.

[3] 陈应虎, 杨哲, 艾传鲜. 数据中心APT攻击检测和防御技术[J]. 网络安全技术与应用,2023(6): 4-7.
CHEN Y H, YANG Z, AI C X. Data center APT attack detection and defense technology[J]. Network Security Technology & Application, 2023 (6): 4-7.

[4] ARULKUMAR D, KARTHEEBAN K, ARULKUMARAN G. The APT cyber warriors with TTP weapons to battle: an review on IoT and cyber twin[M]. Pennsylvania: IGI Global, 2022.

[5] 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.

[6] TATAM M, SHANMUGAM B, AZAM S, et al. A review of threat modelling approaches for APT-style attacks[J]. Heliyon, 2021, 7(1): e05969.

[7] TALIB A M, NASIR Q, BOU NASSIF A, et al. APT beaconing detection: a systematic review[J]. Computers & Security, 2022, 122: 102875.

[8] ALSHAMRANI A, MYNENI S, CHOWDHARY A, et al. A survey on advanced persistent threats: techniques, solutions, challenges, and research opportunities[J]. IEEE Communications Surveys & Tutorials, 2019, 21(2): 1851-1877.

[9] SHARMA A, GUPTA B B, SINGH A K, et al. Advanced persistent

threats (APT): evolution, anatomy, attribution and countermeasures[J]. Journal of Ambient Intelligence and Humanized Computing, 2023, 14 (7): 9355-9381.

[10] AUTY M. Anatomy of an advanced persistent threat[J]. Network Security, 2015, 2015(4): 13-16.

[11] 张瑜, 潘小明, LIU Qingzhong, 等. APT攻击与防御[J]. 清华大学学报(自然科学版), 2017, 57(11): 1127-1133.
ZHANG Y, PAN X M, LIU Q Z, et al. APT attacks and defenses[J]. Journal of Tsinghua University (Science and Technology), 2017, 57 (11): 1127-1133.

[12] NIST. Managing information security risk: organization, mission, and information system view[R]. US Department of Commerce, 2011.

[13] CALTAGIRONE S, PENDERGAST A, BETZ C. The diamond model of intrusion analysis[J]. Threat Connect, 2013, 298(704): 1-61.

[14] HUTCHINS E, CLOPPERT M, AMIN R. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains[J] Leading Issues in Information Warfare & Security Research, 2011, 1(1): 80-106.

[15] LEVY E. Poisoning the software supply chain[J]. IEEE Security & Privacy, 2003, 1(3): 70-73.

[16] STROM B E, APPLEBAUM A, MILLER D P, et al. Mitre attack: design and philosophy[M]. Virginia : MITRE Corporation, 2018.

[17] CENTER M I. APT1: exposing one of China's cyber espionage units [R]. 2013.

[18] KHALEEFA E J, ABDULAH D A, Concept and difficulties of advanced persistent threats (APT): survey[J]. International Journal of Nonlinear Analysis and Applications, 2022, 13(1): 4037-4052.

[19] China Information Technology Security Evaluation Center. Global advanced persistent threat (APT) research report[R]. 2023 .

[20] MAZURCZYK W, CAVIGLIONE L. Cyber reconnaissance techniques[J]. Communications of the ACM, 2021, 64(3): 86-95.

[21] ROY S, SHARMIN N, ACOSTA J C, et al. Survey and taxonomy of adversarial reconnaissance techniques[J]. ACM Computing Surveys, 2023, 55(6): 1-38.

[22] SHAIKH S A, CHIVERS H, NOBLES P, et al. Network reconnaissance [J]. Network Security, 2008(11): 12-16.

[23] GONT F, CHOWN T. Network reconnaissance in IPv6 networks[J]. RFC, 2016, 7707: 1-38.

[24] BOU-HARB E, DEBBABI M, ASSI C. Cyber scanning: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2014, 16 (3): 1496-1519.

[25] SALAHDINE F, KAABOUCH N. Social engineering attacks: a survey[J]. Future Internet, 2019, 11(4): 89.

[26] KANTA A, COISEL I, SCANLON M. A survey exploring open source Intelligence for smarter password cracking[J]. Forensic Science International: Digital Investigation, 2020, 35: 301075.

[27] RAI B K, VERMA R, TIWARI S. Using open source intelligence as a tool for reliable web searching[J]. SN Computer Science, 2021, 2 (5): 402.

[28] MILLAR K A. Graph-based machine learning for passive network reconnaissance within encrypted networks[D]. Australia: The University of Adelaide, 2022.

[29] USSATH M, JAEGER D, CHENG F, et al. Advanced persistent threats: Behind the scenes[C]//Proceedings of the 2016 Annual Conference on Information Science and Systems (CISS). Piscataway: IEEE Press, 2016: 181-186.

[30] FireEye Company. HAMMERTOSS: stealthy tactics define a Russian cyber threat group[R]. 2015.

[31] FIFIELD D, LAN C, HYNES R, et al. Blocking-resistant communication through domain fronting[J]. Proceedings on Privacy Enhancing Technologies, 2015(2): 46-64.

[32] CHARI S, HALEVI S, VENEMA W. Where do you want to go today? escalating privileges by pathname manipulation[C]//Proceedings of the Symposium on Network and Distributed System Security. Piscataway: IEEE Press, 2010: 1-16.

[33] DAVI L, DMITRIENKO A, SADEGHI A R, et al. Privilege escalation attacks on android[C]//International Conference on Information Security. Berlin: Springer, 2011: 346-360.

[34] SUCIU D, MCLAUGHLIN S, SIMON L, et al. Horizontal privilege escalation in trusted applications[C]//Proceeding of the 29th USENIX Security Symposium. Berkeley: USENIX Association, 2020: 825-840.

[35] KIM Y M, LEE B. Extending a hand to attackers: browser privilege escalation attacks via extensions[C]//Proceeding of the 32nd USENIX Security Symposium. Berkeley: USENIX Association, 2023: 7055-7071.

[36] Red Hat Customer Portal. Kernel local privilege escalation "dirty COW"[R]. 2016.

[37] SALEEL A P, NAZEER M, BEHESHTI B D. Linux kernel OS local root exploit[C]//Proceedings of the 2017 IEEE Long Island Systems, Applications and Technology Conference (LISAT). Piscataway: IEEE Press, 2017: 1-5.

[38] LIN Z P, WU Y H, XING X Y. DirtyCred: escalating privilege in linux kernel[C]//Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2022: 1963-1976.

[39] HABER M J, ROLLS D. A nuance on lateral movement[M]. Berkeley: Apress, 2020.

[40] DEMERS D, LEE H. Kerberoasting: case studies of an attack on a cryptographic authentication technology[J]. International Journal of Cybersecurity Intelligence & Cybercrime, 2022, 5(2): 25-39.

[41] GARCIA L A, BRASSER F, CINTUGLU M H, et al. Hey, my malware knows physics! attacking PLCs with physical model aware rootkit[C]// Proceedings of 2017 Network and Distributed System Security Symposium. Reston: Internet Society, 2017: 1-15.

[42] WAMPLER J, MARTINY I, WUSTROW E. ExSpectre: hiding malware in speculative execution[C]//Proceedings of the 2019 Network and Distributed System Security Symposium. Reston: Internet Society, 2019.

[43] ULLAH F, EDWARDS M, RAMDHANY R, et al. Data exfiltration: a review of external attack vectors and countermeasures[J]. Journal of Network and Computer Applications, 2018, 101: 18-54.

[44] NAR K, SASTRY S S. An analytical framework to address the data exfiltration of advanced persistent threats[C]//Proceedings of the 2018 IEEE Conference on Decision and Control (CDC). Piscataway: IEEE Press, 2018: 867-873.

[45] CHIEN E, OMURCHU L, FALLIERE N. W32. Duqu: the precursor to the next stuxnet[C]//Proceeding of the 5th USENIX Workshop on Large-Scale Exploits and Emergent Threats (LEET 12). Berkeley: USENIX Association, 2012: 5.

[46] HALFOND W G, VIEGAS J, ORSO A. A classification of SQL-injection attacks and countermeasures[C]//Proceedings of the IEEE In-

ternational Symposium on Secure Software Engineering. Piscataway: IEEE Press, 2006: 13-15.

[47] CLARKE J. Blind SQL injection exploitation[M]. Amsterdam: Elsevier, 2009.

[48] MUKHERJEE S, SEN P, BORA S, et al. SQL Injection: a sample review[C]//Proceedings of the 2015 6th International Conference on Computing, Communication and Networking Technologies (ICCCNT). Piscataway: IEEE Press, 2015: 1-7.

[49] BAKLIZI M, ATOUM I, ABDULLAH N, et al. A technical review of SQL injection tools and methods: a case study of SQLMap[J]. International Journal of Intelligent Systems and Applications in Engineering, 2022, 10(3): 75-85.

[50] ANU P, VIMALA S. A survey on sniffing attacks on computer networks [C]//Proceedings of the 2017 International Conference on Intelligent Computing and Control (I2C2). Piscataway: IEEE Press, 2017: 1-5.

[51] GLĂVAN D. Sniffing attacks on computer networks[J]. Scientific Bulletin of Naval Academy, 2020, 23(1): 202-207.

[52] D'ORAZIO C J, CHOO K K R, YANG L T. Data exfiltration from Internet of things devices: iOS devices as case studies[J]. IEEE Internet of Things Journal, 2017, 4(2): 524-535.

[53] SARKAR S, LIU J Q, JOVANOV E. A robust algorithm for sniffing BLE long-lived connections in real-time[C]//Proceedings of the 2019 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2019: 1-6.

[54] PANJWANI S, TAN S, JARRIN K M, et al. An experimental evaluation to determine if port scans are precursors to an attack[C]//Proceedings of the 2005 International Conference on Dependable Systems and Networks (DSN'05). Piscataway: IEEE Press, 2005: 602-611.

[55] ANDERSON J P. Computer security threat monitoring and surveillance[M]. Washington: James P. Anderson Company, 1980.

[56] DENNING D E. An intrusion-detection model[J]. IEEE Transactions on Software Engineering, 1987, SE-13(2): 222-232.

[57] KUMAR V, SANGWAN O P. Signature based intrusion detection system using SNORT[J]. International Journal of Computer Applications & Information Technology, 2012, 1(3): 35-41.

[58] HEBERLEIN L T, DIAS G V, LEVITT K N, et al. A network security monitor[C]//Proceedings of the 1990 IEEE Computer Society Symposium on Research in Security and Privacy. Piscataway: IEEE Press, 1990: 296-304.

[59] KIM J, LEE J H. A slow port scan attack detection mechanism based on fuzzy logic and a stepwise policy[C]//Proceedings of the 2008 IET 4th International Conference on Intelligent Environments. London: IET, 2008: 1-5.

[60] CONTI G, ABDULLAH K. Passive visual fingerprinting of network attack tools[C]//Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security. New York: ACM Press, 2004: 45-54.

[61] CHEN J J, CHENG X J. A novel fast port scan method using parthenogenetic algorithm[C]//Proceedings of the 2009 2nd IEEE International Conference on Computer Science and Information Technology. Piscataway: IEEE Press, 2009: 219-222.

[62] ALMOMANI O, ALMAIAH M A, MADI M, et al. Reconnaissance attack detection via boosting machine learning classifiers[C]//Proceedings of the 4th International Computer Sciences and Informatics Conference (ICSIC 2022). New York: ACM Press, 2023.

[63] CROUSE M, PROSSER B, FULP E W. Probabilistic performance analysis of moving target and deception reconnaissance defenses[C]// Proceedings of the 2nd ACM Workshop on Moving Target Defense. New York: ACM Press, 2015: 21-29.

[64] FERGUSON-WALTER K J, MAJOR M M, JOHNSON C K, et al. Examining the efficacy of decoy-based and psychological cyber deception [C]//Proceedings of the 30th USENIX security symposium (USENIX Security 21). Berkeley: USENIX Association, 2021: 1127-1144.

[65] ZHONG Z X, FAN W J. A honey-imprint enabled approach for resisting social engineering attacks[C]//Proceedings of the 2023 24st Asia-Pacific Network Operations and Management Symposium (APNOMS). Piscataway: IEEE Press, 2023: 302-305.

[66] YU X F, SAMARASINGHE N, MANNAN M, et al. Got sick and tracked: privacy analysis of hospital websites[C]//Proceedings of the 2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). Piscataway: IEEE Press, 2022: 278-286.

[67] RUGO A, ARDAGNA C A. Transparency-based reconnaissance for APT attacks[C]//Proceedings of the 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC). Piscataway: IEEE Press, 2023: 1652-1657.

[68] TSINGANOS N, MAVRIDIS I, GRITZALIS D. Utilizing convolutional neural networks and word embeddings for early-stage recognition of persuasion in chat-based social engineering attacks[J]. IEEE Access, 2022, 10: 108517-108529.

[69] PEIRETTI F, PENSA R G. Detection of privacy-harming social media posts in Italian[C]//International Symposium on Security and Privacy in Social Networks and Big Data. Berlin: Springer, 2023: 203-223.

[70] RAJESWARY C, THIRUMARAN M. A comprehensive survey of automated website phishing detection techniques: a perspective of artificial intelligence and human behaviors[C]//Proceedings of the 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS). Piscataway: IEEE Press, 2023: 420-427.

[71] KIM T, PARK N, HONG J, et al. Phishing URL detection: a network-based approach robust to evasion[C]//Proceedings of the Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2022: 1769-1782.

[72] HAFNER L, WUTZ F, PÖHN D, et al. TASEP: a collaborative social engineering tabletop role-playing game to prevent successful social engineering attacks[C]//Proceedings of the 18th International Conference on Availability, Reliability and Security. New York: ACM Press, 2023: 1-10.

[73] DUMAN S, KALKAN-CAKMAKCI K, EGELE M, et al. EmailProfiler: spearphishing filtering with header and stylometric features of emails[C]//Proceedings of the 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC). Piscataway: IEEE Press, 2016: 408-416.

[74] CIDON A, GAVISH L, BLEIER I, et al. High precision detection of business email compromise[C]//Proceeding of the 28th USENIX Security Symposium. Berkeley: USENIX Association, 2019: 1291-1307.

[75] BAI B, FENG Y, LIU B X, et al. APT behaviors detection based on email business scenarios[C]//Proceedings of the 2021 IEEE Sixth International Conference on Data Science in Cyberspace (DSC). Piscataway: IEEE Press, 2021: 171-178.

[76] ALLEN J, YANG Z, LANDEN M, et al. Mnemosyne: an effective and efficient postmortem watering hole attack investigation system[C]//Pro-

ceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2020: 787-802.

[77] ASLAN Ö A, SAMET R. A comprehensive review on malware detection approaches[J]. IEEE Access, 2020, 8: 6249-6271.

[78] WANG X, ZHENG K F, NIU X X, et al. Detection of command and control in advanced persistent threat based on independent access[C]//Proceedings of the 2016 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2016: 1-6.

[79] YAN G H, LI Q, GUO D, et al. Discovering suspicious APT behaviors by analyzing DNS activities[J]. Sensors, 2020, 20(3): 731.

[80] MENON A. Thwarting C2 communication of DGA-based malware using process-level DNS traffic tracking[C]//Proceedings of the 2019 7th International Symposium on Digital Forensics and Security (ISDFS). Piscataway: IEEE Press, 2019: 1-5.

[81] TRAN H, DANG C, NGUYEN H, et al. Multi-confirmations and DNS graph mining for malicious domain detection[C]//Proceedings of the Intelligent Computing Conference. Berlin: Springer, 2019: 639-653.

[82] HIGHNAM K, PUZIO D, LUO S, et al. Real-time detection of dictionary DGA network traffic using deep learning[J]. SN Computer Science, 2021, 2(2): 110.

[83] YUN X C, XIE J, LI S H, et al. Detecting unknown HTTP-based malicious communication behavior via generated adversarial flows and hierarchical traffic features[J]. Computers & Security, 2022, 121: 102834.

[84] YAMAUCHI T, AKAO Y, YOSHITANI R, et al. Additional kernel observer to prevent privilege escalation attacks by focusing on system call privilege changes[C]//Proceedings of the 2018 IEEE Conference on Dependable and Secure Computing (DSC). Piscataway: IEEE Press, 2018: 1-8.

[85] ELSABAGH M, JOHNSON R, STAVROU A, et al. FIRMSCOPE: automatic uncovering of privilege-escalation vulnerabilities in pre-installed apps in android firmware[C]//Proceedings of the 29th USENIX security symposium. Berkeley: USENIX Association. 2020: 2379-2396.

[86] LI R, DIAO W R, LI Z, et al. Android custom permissions demystified: from privilege escalation to design shortcomings[C]//Proceedings of the 2021 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2021: 70-86.

[87] TIAN Z H, SHI W, WANG Y H, et al. Real-time lateral movement detection based on evidence reasoning network for edge computing environment[J]. IEEE Transactions on Industrial Informatics, 2019, 15(7): 4285-4294.

[88] YAN D Y, LIU F, JIA K. Modeling an information-based advanced persistent threat attack on the internal network[C]//Proceedings of the 2019 IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2019: 1-7.

[89] MOHAMED N, BELATON B. SBI model for the detection of advanced persistent threat based on strange behavior of using credential dumping technique[J]. IEEE Access, 2021, 9: 42919-42932.

[90] BAI T, BIAN H B, ABOU DAYA A, et al. A machine learning approach for RDP-based lateral movement detection[C]//Proceedings of the 2019 IEEE 44th Conference on Local Computer Networks (LCN). Piscataway: IEEE Press, 2019: 242-245.

[91] BI J C, HE S B, LUO F J, et al. Defense of advanced persistent threat on industrial Internet of Things with lateral movement modeling[J]. IEEE Transactions on Industrial Informatics, 2023, 19(9): 9619-9630.

[92] PACCAGNELLA R, LIAO K, TIAN D, et al. Logging to the danger zone: race condition attacks and defenses on system audit frameworks [C]//Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2020: 1551-1574.

[93] PACCAGNELLA R, DATTA P, HASSAN W U, et al. Custos: practical tamper-evident auditing of operating systems using trusted execution [C]//Proceedings of the 2020 Network and Distributed System Security Symposium. Reston: Internet Society, 2020.

[94] JIANG P, HUANG R Z, LI D, et al. Auditing frameworks need resource isolation: a systematic study on the super producer threat to system auditing and its mitigation[C]//Proceedings of the 32nd USENIX Security Symposium. Berkeley: USENIX Association. 2023: 355-372.

[95] SEKAR R, KIMM H, AICH R, et al. eAudit: a fast, scalable and deployable audit data collection system[C]//Proceedings of the 2024 IEEE Symposium on Security and Privacy (S&P). Piscataway: IEEE Press, 2023: 87-87.

[96] TIAN D H, MA R, JIA X Q, et al. A kernel rootkit detection approach based on virtualization and machine learning[J]. IEEE Access, 2019, 7: 91657-91666.

[97] NAGY R, NÉMETH K, PAPP D, et al. Rootkit detection on embedded IoT devices[J]. Acta Cybernetica, 2021, 25(2): 369-400.

[98] YAO M X, FULLER J, KASTURI R P, et al. Hiding in plain sight: an empirical study of web application abuse in malware[C]//Proceedings of the 32nd USENIX Security Symposium. Berkeley: USENIX Association, 2023: 6115-6132.

[99] LAL A, PRASAD A, KUMAR A, et al. Data exfiltration: preventive and detective countermeasures[C]//Proceedings of the International Conference on Innovative Computing & Communication (ICICC). Amsterdam: Elsevier, 2022: 1-8.

[100] KIPERBERG M, AMIT G, YESHOOROON A, et al. Efficient DLPvisor: an efficient hypervisor-based DLP[C]//Proceedings of the 2021 IEEE/ACM 21st International Symposium on Cluster, Cloud and Internet Computing (CCGrid). Piscataway: IEEE Press, 2021: 344-355.

[101] ALKILANI H, NASEREDDIN M, HADI A, et al. Data exfiltration techniques and data loss prevention system[C]//Proceedings of the 2019 International Arab Conference on Information Technology (ACIT). Piscataway: IEEE Press, 2019: 124-127.

[102] ALMOMANI I, ALKHAYER A, EL-SHAFAI W. A crypto-steganography approach for hiding ransomware within HEVC streams in android IoT devices[J]. Sensors, 2022, 22(6): 2281.

[103] FATHI-KAZEROONI S, ROJAS-CESSA R. GAN tunnel: network traffic steganography by using GANs to counter Internet traffic classifiers[J]. IEEE Access, 2020, 8: 125345-125359.

[104] PŁACHTA M, KRZEMIEŃ M, SZCZYPIORSKI K, et al. Detection of image steganography using deep learning and ensemble classifiers[J]. Electronics, 2022, 11(10): 1565.

[105] ZHAI L M, WANG L N, REN Y Z. Universal detection of video steganography in multiple domains based on the consistency of motion vectors[J]. IEEE Transactions on Information Forensics and Security, 2019, 15: 1762-1777.

[106] HU Y T, HUANG Y H, YANG Z L, et al. Detection of heterogeneous parallel steganography for low bit-rate VoIP speech streams[J]. Neurocomputing, 2021, 419: 70-79.

[107] NADLER A, AMINOV A, SHABTAI A. Detection of malicious and

low throughput data exfiltration over the DNS protocol[J]. Computers & Security, 2019, 80: 36-53.

[108] CRESPO-MARTINEZ I S, CAMPAZAS-VEGA A, GUERRERO-HIGUERAS Á M, et al. SQL injection attack detection in network flow data[J]. Computers & Security, 2023, 127: 103093.

[109] NYAKOMITTA P S, ABEKA S O. A Survey of data exfiltration prevention techniques[J]. International Journal of Scientific Research in Science and Technology, 2020, 5(8): 8.

[110] WEN J, TAN C H, CHEN J N, et al. The application of RNN-based API data security detection in government cloud service[C]//Proceedings of the 2022 Tenth International Conference on Advanced Cloud and Big Data (CBD). Piscataway: IEEE Press, 2022: 276-281.

[111] 郭嘉琰, 李荣华, 张岩, 等. 基于图神经网络的动态网络异常检测算法[J]. 软件学报, 2020, 31(3): 748-762.
GUO J Y, LI R H, ZHANG Y, et al. Graph neural network based anomaly detection in dynamic networks[J]. Journal of Software, 2020, 31(3): 748-762.

[112] CHUNG M H, CHIGNELL M, WANG L, et al. Interactive machine learning for data exfiltration detection: active learning with human expertise[C]//Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Piscataway: IEEE Press, 2020: 280-287.

[113] MARQUES R S, EPIPHANIOU G, AL-KHATEEB H, et al. A flow-based multi-agent data exfiltration detection architecture for ultra-low latency networks[J]. ACM Transactions on Internet Technology, 2021, 21(4): 1-30.

[114] DIJK A. Detection of advanced persistent threats using artificial intelligence for deep packet inspection[C]//Proceedings of the 2021 IEEE International Conference on Big Data (Big Data). Piscataway: IEEE Press, 2021: 2092-2097.

[115] HOSSAIN M N, MILAJERDI S M, WANG J N, et al. SLEUTH: real-time attack scenario reconstruction from COTS audit data[J]. arXiv Preprint, arXiv: 1801.02062, 2018.

[116] MILAJERDI S M, ESHETE B, GJOMEMO R, et al. POIROT: aligning attack behavior with kernel audit records for cyber threat hunting [C]//Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security. New York: ACM Press, 2019: 1798-1812.

[117] MILAJERDI S M, GJOMEMO R, ESHETE B, et al. HOLMES: real-time APT detection through correlation of suspicious information flows[C]//Proceedings of the 2019 IEEE Symposium on Security and Privacy (SP). Piscataway: IEEE Press, 2019: 1137-1152.

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

[119] ZHAO J, YAN Q B, LIU X D, et al. Cyber threat intelligence modeling based on heterogeneous graph convolutional network[C]//Proceedings of the 23rd International Symposium on Research in Attacks, Intrusions and Defenses. Piscataway: IEEE Press, 2020: 241-256.

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

[121] ALSAHEEL A, NAN Y, MA S, et al. ATLAS: a sequence-based learning approach for attack investigation[C]//USENIX Security Symposium. Berkeley: USENIX Association, 2021: 3005-3022.

[122] ANJUM M M, IQBAL S, HAMELIN B. ANUBIS: a provenance graph-based framework for advanced persistent threat detection[C]//Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing. New York: ACM Press, 2022: 1684-1693.

[123] MAHMOUD M, MANNAN M, YOUSSEF A. APTHunter: detecting advanced persistent threats in early stages[J]. Digital Threats: Research and Practice, 2023, 4(1): 1-31.

[124] JIA S S, XU Y B. The APT detection method based on attack tree for SDN[C]//Proceedings of the 2nd International Conference on Cryptography, Security and Privacy. New York: ACM Press, 2018: 116-121.

[125] KUMAR R, KELA R, SINGH S, et al. APT attacks on industrial control systems: a tale of three incidents[J]. International Journal of Critical Infrastructure Protection, 2022, 37: 100521.

[126] YI S Y, SINGH M M, SODHY G C, et al. Fingerprinting generation for advanced persistent threats (APT) detection using Machine Learning techniques[C]//Proceedings of the 2023 13th International Conference on Information Technology in Asia (CITA). Piscataway: IEEE Press, 2023: 31-36.

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

[128] JI Z J, CHOI E, GAO P. A knowledge base question answering system for cyber threat knowledge acquisition[C]//Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2022: 3158-3161.

[129] GOLDREICH O. Secure multi-party computation[J]. Manuscript, 1998, 78(1): 86-97.

[130] FAN J F, VERCAUTEREN F. Somewhat practical fully homomorphic encryption[J]. IACR Cryptol EPrint Arch, 2012, 144: 1-19.

[131] MAVROVOUNIOTIS S, GANLEY M. Hardware security modules [M]. Berlin: Springer, 2014.

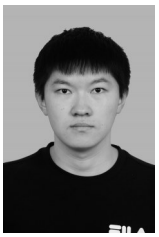
[132] LEE G, SHIM S, CHO B, et al. Fileless cyberattacks: Analysis and classification[J]. ETRI Journal, 2021, 43(2): 332-343.

[作者简介]

王郅伟（2000– ），男，河南信阳人，中国科学院大学博士生，主要研究方向为Fuzzing、APT攻击与防御技术等。

何睎杰（2001– ），女，云南曲靖人，中国科学院大学博士生，主要研究方向为信息安全等。

易鑫（2001– ），男，四川南充人，中国科学院大学硕士生，主要研究方向为信息安全等。



李孜旸（2002– ），男，云南曲靖人，中国科学院大学博士生，主要研究方向为信息安全等。

曹旭栋（1997– ），男，陕西渭南人，中国科学院大学博士生，主要研究方向为网络与系统安全。

尹涛（1989– ），男，重庆人，博士，中关村实验室高级工程师，主要研究方向为网络威胁检测与溯源。

李书豪（1983– ），男，山西文水人，博士，中关村实验室正高级工程师、博士生导师，主要研究方向为网络威胁检测与溯源、人工智能与网络安全等。

付安民（1981– ），男，湖北通城人，博士，南京理工大学教授、博士生导师，主要研究方向为物联网安全、密码学和隐私保护等。

张玉清（1966– ），男，陕西宝鸡人，博士，中国科学院大学教授、博士生导师，主要研究方向为网络攻防与系统安全、大数据与智能安全、物联网系统安全。