网络资产探测技术研究

王宸东 郭渊博 甄帅辉 杨威超

(战略支援部队信息工程大学 郑州 450001) (数学工程与先进计算国家重点实验室 郑州 450001)

摘 要 随着网络技术的迅速普及,大量多样化的网络资产为人们的生产、生活提供了极大便利,同时也对其自身的安全管理提出了挑战。准确、全面地进行网络资产探测是实现网络资产有效管理的前提,也是进行威胁分析的基础。首先回顾了网络资产探测的起源与发展历程;然后全面分析了当前常见的 3 种新型网络资产探测方法(主动、被动和基于搜索引擎)及其关键技术,归纳了它们各自的特点;最后,探讨了该技术未来可能的研究方向。

关键词 资产发现,搜索引擎,指纹识别,网络扫描

中图法分类号 TP393.0 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.12.004

Research on Network Asset Detection Technology

WANG Chen-dong GUO Yuan-bo ZHEN Shuai-hui YANG Wei-chao

(Strategic Support Force University of Information Engineering, Zhengzhou 450001, China)

(State Key Laboratory of Mathematical Engineering and Advanced Computing, Zhengzhou 450001, China)

Abstract With the rapid spread of network technology, large numbers of diversified network assets bring great convenience to people's daily life, but challenges are also posed to their own safety management at the same time. Accurate and comprehensive network asset detection is the prerequisite for the effective management of network assets and the basis for threat analysis. First, this paper reviewed the origin and development process of network asset detection. Next, this paper comprehensively analyzed three common novel methods of network asset detection (active, passive and search engine based) and each key technologies, and summarized the characteristics of these methods respectively. Finally, this paper discussed the development trends and further research directions of this technology.

Keywords Asset detection, Search engine, Fingerprint identification, Network scanning

1 引言

从 20 世纪 60 年代美国的 ARPANet 到今天的国际互联网,网络技术得到了迅猛发展,越来越多的组织和个人接入互联网。包括网络终端、网络设备、网络服务等在内的网络资产已被广泛应用于各类企、事业单位的日常业务工作,极大地提高了工作效率,促进了业务工作的发展,但也带来了许多问题和隐患。随着单位网络规模的不断扩大,网络资产及其所包含的漏洞类型不断增多,给单位网络安全管理带来了巨大压力。

ISO 13335-1:2004《IT 安全管理指南》中将“任何对组织有价值的东西”定义为资产[1],资产作为 IT 安全管理的对象,包括信息(或数据)、硬件、软件、资金、服务、人员等。Sanders 认为资产不仅包括易量化的服务器和网络设备,还包括数据、人员、流程、知识产权和声誉等[2]。本文主要探讨上述资产中具有网络连接的终端、设备、服务等网络资产。网络资产探测是指追踪、掌握网络资产情况的过程,通常包括主机发现、操作系统识别、服务识别等,是实现网络安全管理的重要前提,在网络安全相关工作中具有广泛的应用价值。

一方面,从网络资产管理的角度看,网络资产探测能够为统一软硬件版本、更新升级软件和设备等工作提供信息基础。通过网络资产探测可以发现旧版本的软件,根据最新的威胁情报准确地启动响应措施,避免其存在的漏洞带来威胁;还可以发现非法资产,为及时分析、处理提供便利,最大限度地降低安全问题带来的损失。

网络资产探测不仅为网络安全监控、威胁态势感知提供了系统认知基础,而且在提高入侵检测系统的效率、安全威胁分析等方面也有较多应用。根据掌握的网络资产情况,可以为入侵检测系统去掉不相关的规则、缩小匹配规则库、提高检测效率,也可对告警信息进行过滤,减轻网络安全管理人员的告警分析压力,把更多精力放在处理有效攻击上[3]。同时,面对日益加剧的新型高级持续性(Advanced Persistent Threat)攻击,大规模网络的安全管理人员可在网络资产探测结果的基础上,综合网络资产、网络拓扑结构、漏洞等信息,基于攻击图技术对可能的高危攻击路径进行评估[4],根据评估结果采

到稿日期:2017-11-30 返修日期:2018-03-19 本文受国家自然科学基金项目(61501515,61602515)资助。

王宸东(1992-),男,硕士生,主要研究方向为网络安全;**郭渊博**(1975-),男,教授,博士生导师,主要研究方向为网络与信息安全,E-mail:yuanbo_g@hotmail.com(通信作者);**甄帅辉**(1987-),男,硕士生,主要研究方向为网络安全;**杨威超**(1991-),男,硕士生,主要研究方向为网络安全。

取重点防御和响应措施，从而提高防御的针对性。

另一方面，从安全渗透测试（或攻击）的角度看，网络资产探测也可用于渗透（或攻击）前的信息收集[5]，通过网络资产探测，了解目标网络内主机的操作系统类型、开放端口及其后所运行的应用程序类型和版本信息。掌握目标网络的安全状况，有助于选取高效的渗透（或攻击）方法。

如图 1 所示，根据网络资产探测技术的发展历程，可将其划分为传统和新型两大类。传统类探测技术从繁琐的人工统计发展到基于客户端的自动统计[6-8]；新型类探测技术从入侵式的低速局域网扫描[9-12]发展到高速大规模网络扫描[13-16]，从单一基于流量分析的被动探测[17-24]发展到基于搜索查询实现的非入侵式探测[25-29]。网络资产探测技术为适应网络认知的现实需求而不断发展。

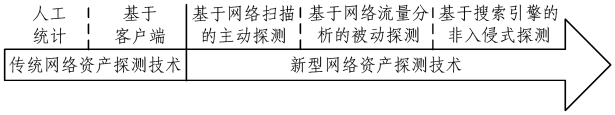

图 1　网络资产探测技术发展历程

Fig. 1　Development process of network asset detection technology

本文首先回顾了网络资产探测技术的起源，从主动、被动和基于搜索引擎这 3 种探测方法出发，归纳梳理了各自的特点与缺陷，分析了新型网络资产探测技术所涉及的高速网络扫描、网络流量分析和网络资产指纹识别技术。然后，讨论了多种探测方法的综合运用、特殊应用场景探测的实现、人工智能技术的结合等未来可能的研究方向。

2　传统网络资产探测

人工统计是最原始的资产探测方法。为了便于资产的管理，通过人工统计定期地组织资产普查，并利用一些软件（如 Excel，Spiceworks 等）进行辅助记录。对于一些小型单位而言，人工统计法是一种经济、便捷且有效的方法，但需要耗费大量的人力资源和时间，时效性差，无法及时发现一些恶意接入的网络资产。

在人工统计的基础上，基于客户端的自动统计方法需要在每台设备上安装客户端，以中心服务器定期查询或客户端定期自动上报代替人工统计，提高了工作效率。但该方法由于需要在所有被探测的网络资产上安装客户端，入侵性最强，在现实工作中可能存在很多限制因素，其探测能力完全取决于客户端安装的全面性及其信息获取能力，且客户端的设计开发需要增加对网络资产多样化操作系统类型的支持，成本较高。

传统网络资产探测方法在资产管理和监控领域也有着广泛的应用，常见的企业网络资产管理解决方案 Spiceworks[6]、美国 IBM 公司的 MAXIMO 系统[7]、加拿大的 Senergy 系统[8]等都采用了上述方法。Spiceworks 是一款适用于中小企业的免费网络资产管理软件，支持人工录入和客户端上报两种探测方式；MAXIMO 和 Senergy 则是基于 C/S 架构的功能更加丰富的商业化网络资产管理软件。

3　新型网络资产探测

随着单位业务的多样化，各类支撑平台和信息管理系统增多，网络规模不断扩大，网络设备、安全设备等越来越复杂，信息安全管理部门同业务部门之间的协调难度也日益增大，传统的以数据库、资产探针为支撑的网络资产探测方法已经难以满足现实需求，一些基于网络扫描、网络流量分析、搜索引擎等技术的新型网络资产探测技术得到了迅速发展。根据探测基础数据来源的不同，新型网络资产探测技术主要分为主动、被动和基于搜索引擎三大类，其基本实现流程如图 2 所示。本节在介绍上述 3 种类型的网络资产探测方法的基础上，对其中涉及的高速网络扫描、网络流量分析、网络资产指纹识别 3 项关键技术进行了深入分析。

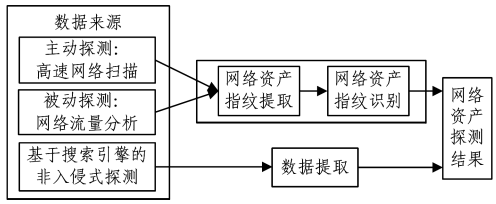

图 2　新型网络资产探测的基本实现流程

Fig. 2　Basic realization process of novel network asset detection

3.1　探测方法

3.1.1　主动探测

主动探测类方法是指通过主动向目标网络资产发送构造的数据包，并从返回数据包的相关信息（包括各层协议内容、包重传时间等）中提取目标指纹，与指纹进库中的指纹进行比对，来实现对开放端口、操作系统、服务及应用类型的探测。根据使用的指纹信息类型，主动探测类方法主要分为基于响应协议栈指纹的主动探测方法和基于单包响应时延统计两类。

Nmap[9]是基于响应协议栈指纹的网络资产探测工具的典型代表，该工具通过向目标资产发送 SYN 包或构造的其他各层协议（IP，UDP，ICMP）数据包，根据应答数据包的特征，进行主机发现和端口、服务、操作系统等的识别。DNmap[13]通过引入分布式的架构提高了探测效率。ZMap[14]、Masscan[15]基于异步无状态扫描工具对扫描机制进行了改进，但通常只能进行端口扫描和主机发现，对操作系统、服务及应用的探测则无能为力，其以弱化功能全面性为代价来增强探测时效性。Yarochkin 等发布的 Xprobe2[10]使用模糊矩阵统计分析探测数据包触发的 ICMP 响应消息特征，以识别目标资产的操作系统类型。

基于响应协议栈指纹的主动探测方法需要发送大量的网络流量，为了减少探测数据包的发送数量，Shamsi 等提出了基于单包响应时延统计理论的操作系统指纹识别工具 Hershel[11]，其只发送单个 SYN 包进行探测，引入随机模型对 SYN/ACK 重传数据包的超时时间（RTO）指纹进行分析和识别。该模型充分考虑了网络抖动延迟、丢包、人为协议栈内容修改等多种因素。为了应对响应时延观测数据中的噪声，Shamsi 等又提出了支持自动化指纹特征生成的模型和算法[12]。

主动探测方法相比于传统方法便捷且高效，其通过目标网络内的一个节点进行探测数据包的收发和响应分析实现，不需要在所有网络资产上安装客户端。但该方法存在以下问

题:1)探测行为所引发的大量网络流量噪声很容易对一些正常运行的系统造成影响[30],因此不适用于需要持续运行的关键性系统,如在线服务的金融系统、工控领域的数据采集和监控系统(SCADA)、电子管理系统(EMS)、过程控制(PCS)系统等[31];2)易触发各类安全设备的警报,不利于渗透测试过程中信息收集行为的隐蔽性;3)对一些受到代理、NAT 路由以及安全设备保护的网络资产的探测难度大;4)探测结果的全面性相对有限,每次探测只能了解该时刻的网络资产状态;5)对一些客户端模式的软件、瞬时的服务、需要特定数据包激活的服务等资产进行探测时无效。

3.1.2 被动探测

被动探测方法是指采集目标网络的流量,对流量中应用层 HTTP,FTP,SMTP 等协议数据包中的特殊字段 banner 或 IP、TCP 三次握手、DHCP 等协议数据包的指纹特征进行分析,从而实现对网络资产信息的被动探测。一般情况下,部分应用层协议数据包中会包含一些网络资产信息,如:HTTP 协议的 User-Agent 字段中包含了操作系统、浏览器版本等信息,但该字段非常容易被修改,可信度不高,而且此分析方式对使用加密协议的数据包无效。因此,目前相关研究更侧重于对应用层以下的 IP,TCP,DHCP 等协议特征进行分析。常见的被动网络资产探测工具有 p0f,PRADS,Satori,Network Miner 等。

p0f[17]是一款基于流量分析的操作系统被动识别工具,通过捕获目标主机发出的数据包分析 TCP 三次握手数据包头、IP 头的特殊字段设置,对主机上的操作系统信息进行鉴别。由于仅通过分析 TCP 握手信息发现操作系统 TCP/IP 协议栈,降低了内存的开销,但其所能提供的信息相对有限。为了应对网络规模增长带来的大量网络流量的处理,Barnes 等提出了在 Linux 内核空间部署的 K-p0f[18],K-p0f 去掉了 p0f 所有与操作系统探测不相关的功能(如 HTTP 客户端发现、物理连接类型发现等),网络流量的处理速度比 p0f 提高了近 16 倍。陈军等提出采用多线程机制的 Multiple-p0f[19],在主线程启动后,创建 pcap 抓包线程、Receive 线程、Write 数据库线程及 p0f 进程池,均使用队列进行顺序处理,从而使网络数据包的处理速度提高了近 4 倍。

PRADS[20](Passive Real-time Asset Detection System)是一款通过被动监听网络流量数据(包括 TCP、UDP、ICMP、ARP 数据包)发现在线主机、操作系统和服务类型的工具,其前身 PADS[21]是基于网络流量特征的服务、应用及其版本识别的网络资产探测工具。PRADS 是 p0f 和 PADS 的集成[22],其操作系统的探测能力来自 p0f,厂商、MAC 地址识别和服务、应用探测能力来自 PADS。PRADS 主要针对使用 TCP 的服务,能够发现在非标准端口的 TCP 服务,但对于 UDP 和 ICMP 流量,仅能探测标准端口上运行的服务。

Satori[23]是基于 DHCP 消息中的选项及其排列顺序特征对网络资产和操作系统类型进行识别的工具,准确性较高,但由于 DHCP 消息仅在局域网内可见,该方法的适用范围有限。Network Miner[24]是 windows 平台下基于 Satori 和 p0f 指纹库的网络取证工具,p0f 和 PRADS 将每个数据包的识别结果直接显示,而该工具更加注重主机的识别,将探测结果按主机进行划分和提取。

被动探测方法通过分析采集的网络流量得到资产信息,对目标网络运行的影响小,无额外网络流量插入;对安全设备保护的网络资产也具备探测能力;便于长期历史数据的积累,从而掌握网络资产发展变化的过程。但探测的全面性和高效性受限于所分析网络流量的全面性。由于需要获取目标网络的大量流量数据作为分析基础,该方法适用的网络规模有限;对探测过程中不在线或不产生网络流量的网络资产无效。

3.1.3 基于搜索引擎的非入侵式探测

随着大数据、云计算等新技术的引入,搜索引擎的功能得到了加强和拓展,远超出了传统的网页检索,类型也日趋多样,除了常见的谷歌、百度、Bing 等通用搜索引擎之外,还出现了 Shodan、Censys 及 ZoomEye 等网络安全专用的搜索引擎。搜索引擎技术的发展为实现基于搜索引擎的非入侵式网络资产探测提供了基础。

如图 3 所示,不同于前文所述的主、被动探测方法需要同目标网络资产交互或者获取网络流量,基于搜索引擎的探测方法能够以一种搜索查询的方式间接、高效地完成大规模网络资产探测。根据使用搜索引擎的类型,可将其分为基于通用搜索引擎和基于网络安全专用搜索引擎两大类[25],下面分别对其进行介绍。

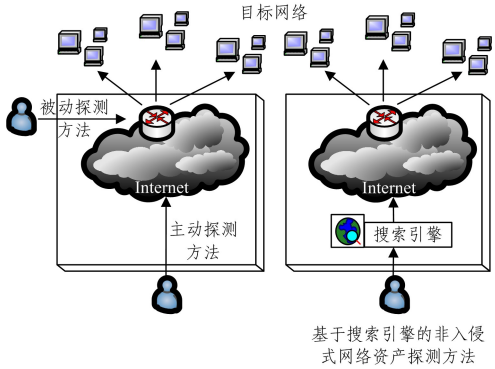

图 3 基于搜索引擎的非入侵式探测与其他探测方法的对比

Fig. 3 Comparison between search engine based non-intrusive detection method and other methods

1)基于通用搜索引擎的探测

谷歌作为全球规模最大的搜索引擎,索引的页面总数已过万亿,为用户提供了搜索、开发等功能多样的服务,是通用搜索引擎的典型代表。在网络安全领域,谷歌搜索引擎也发挥着重要作用。谷歌黑客技术是一种利用谷歌搜索引擎进行漏洞目标探测以及敏感信息挖掘的技术,可以实现网站映射、查看站点目录列表、查找登录页面、查找口令文件、查找网络设备等功能,因此具备一定的网络资产探测能力。

GHDB(Google Hacking Database)是一个谷歌黑客搜索查询指令的数据库[5]。黄超[26]提出了基于 GHDB 中的特定搜索查询串、某些服务的页面脚注、Web 服务器返回的错误消息中携带的信息实现端口、操作系统及版本的探测。但由于通用搜索引擎局限于所使用网络爬虫的数据获取范围,基于通用搜索引擎探测方法的探测对象只能以 Web 相关的网络资产为主。

2）基于网络安全专用搜索引擎的探测

2009 年，Matherly 创建了 Shodan 搜索引擎[27]，不同于谷歌、Bing、百度等基于网络爬虫的通用网页搜索引擎，Shodan 侧重于对所有连接互联网的设备及其组件类型信息的搜索。用户可以使用 Shodan 搜索摄像头、打印机，甚至是粒子加速器、核电站控制设备。根据 Shodan 官方说明书[32]，分布在全球 8 个国家和地区的探测器持续不断地运行，实时更新数据库。

2015 年，Durumeric 等将 ZMap 同谷歌云平台相结合，开发了 Censys 系统[28]，该系统使用了基于 Go 语言开发的 ZGrab 应用扫描器与 ZMap 配合，而后进行数据处理汇总，提取结构化数据，并将其保存于谷歌云存储平台；利用开源的 ElasticSearch 平台和谷歌 BigQuery 分别在前端和后台为用户提供 ZMap 全网端口、服务扫描结果的搜索查询。

国内的相关研究有知道创宇公司开发的 ZoomEye 搜索引擎[29]，该公司在 Nmap 的基础上，开发了 Web 指纹识别引擎 Wmap，并依托其后台的大数据存储处理平台，为用户提供了设备指纹、Web 服务等的搜索功能，并于 2015 年上线了工控专题，支持 12 种工控协议的数据检索[33]。类似的用于网络探测结果查询的搜索引擎还有东北大学谛听（开放式）、傻蛋（开放式）、中电网络信息安全公司工控系统接入互联网威胁感知系统（内部）、绿盟广谱平台（内部）等。

上述基于两类搜索引擎的非入侵式网络资产探测方法，依托从搜索引擎获取的网络爬虫爬取结果或专用服务器扫描结果，为用户提供了一种间接查询实现网络资产探测的方式，这种方式不仅高速、隐蔽，避免了同目标网络的直接交互，为安全管理员审视本单位（或组织）网络资产的安全情况提供了一种新的视角，而且为全网范围的探测和历史数据的积累提供了支持。但它也存在数据获取能力局限于所使用的搜索引擎、对无公网 IP 资产无效等缺陷，且易受到欺骗，准确率相比主动、被动探测而言更低。

表 1 对上述传统方法和 3 种新型网络资产探测方法的主要特点及存在的问题进行了简要归纳。

表 1　现有网络资产探测方法及其特点

Table 1　Current network asset detection methods and their characteristics

类型		范围	主要特点	存在的问题
传统	人工统计	内网，小规模	可以发现新型探测方法无法分析到的部分（不产生网络流量或探测数据包无法触及的网络资产）	耗时费力；时效性差
	基于客户端		需要大规模安装客户端，由客户端自动化采集、上报网络资产数据，速度快，效率高，节省了人力	入侵性最强，限制因素多；客户端开发、设计成本高
新型	主动探测	全网/内网，各种规模均适用	无需安装客户端，在网内一个节点运行并收发探测数据包即可；速度快，对不产生网络流量的资产也能及时发现	噪声大，易触发报警；全面性有限，仅能了解当次探测的状态；对安全设备保护的网络资产探测的难度大
	被动探测	仅限于内网	无网络流量插入，入侵性小；对安全设备保护的网络资产也具备一定的探测能力；支持历史数据的积累	需要获取目标网络的流量数据，适用范围限于内网；探测结果受限于所分析网络流量的全面性；对不产生网络流量数据的资产无效
	搜索引擎 通用网络安全专用	仅限于公网（目标资产必须具有公网 IP）	以查询的方式探测，隐蔽性强，探测速度快；支持全网探测；支持历史数据的积累	仅对 Web 相关网络资产有效 对公网网络组件、网络设备、网络服务等的探测具有优势	无法对内网资产进行探测；受限于搜索引擎的数据获取能力；易被欺骗，准确率相对较低

3.2　关键技术分析

3.2.1　高速网络扫描技术

主动探测方法主要采用的是网络扫描技术。网络扫描是指根据对方服务所采用的协议，在一定时间内，通过自身系统对对方协议进行特定读取、猜想验证，甚至恶意破坏，并将对方直接或间接返回的数据作为某指标判断依据的一种行为[34]。自诞生以来，网络扫描技术在攻防双方的博弈中，从低速的局域网扫描到高速的全网扫描，得到了长足的发展。本节先简要回顾网络扫描技术的发展，然后从扫描机制、扫描地址生成、并行化加速 3 个方面对高速网络扫描技术的特点进行分析。

自 1992 年 Chris 开发早期的扫描工具 Internet Security Scanner（ISS）以来，为 Unix 设计的基于 HTML 界面的 SATAN、生成报告功能更加强大的商用产品 CyberCop Scanner、基于 NASL 安全脚本描述语言可扩展插件的开源工具 Nessus 等网络扫描工具相继出现。1997 年，Lyon 发布了灵活性好且应用最广泛的扫描工具 Nmap[9]。2009 年，Garcia 利用 Python 下的 Twisted 框架开发了 DNmap[13]，DNmap 使用标准的客户端/服务端（C/S）架构创建分布式的 Nmap 扫描网络，以提高网络扫描的速度。

在扫描机制方面，传统 TCP SYN 扫描中，扫描器通常需要建立完整的 TCP 会话，为完成三次握手过程，CPU 需要为其划分专门的数据区，如果探测整个 IPv4 地址空间，将耗费大量的系统资源。高速网络扫描技术则不完成三次握手，只发送第一个 SYN，而后 RST 取消连接，并对该次探测地址进行 Hash，将值保存在缓存中，对方回复的 SYN-ACK 则由专门的接收模块负责等待接收。这种无状态保持的设计，可能会因网络原因丢失一定比例的数据，从实验结果看，这个比例仅占 2%左右[14]，但极大地减少了状态记录的开销。ZMap 使用商用硬件和 Gbit 级网络速率，能在 45 min 内完成整个 IPv4 地址空间的扫描，比传统 Nmap 最激进的默认条件设置下都快 1300 倍[14]。类似的工具 Masscan 使用双端口 10Gbit 级的网卡则仅用 3 min 就能完成全网扫描[15]。

在扫描地址生成方面，新型高速网络扫描工具使用了基于素数域原根[14]（ZMap）或加密算法[15]（Masscan）的地址生成策略，增加了相邻扫描地址的随机性，减少了扫描对同一 IP 地址段内目标网络的压力，不仅实现了高效的资源（带宽、计算）利用，而且对扫描行为起到了一定的隐蔽作用。

在并行化加速方面，ZMap，Masscan 以及 Nmap 的集群式改进版 DNmap 均支持多主机分布式扫描[35]。2014 年，

Adrian 等在 ZMap 的基础上开发的 Zippier ZMap[16] 对基于素数域原根的扫描地址生成进行了分片处理，不同分片的扫描地址集合都是迭代生成的互为不相关的子集，利于地址生成策略的并行化执行，同时其结合了黑名单数据结构、数据包传输机制等其他改进，适应了 10Gbps 环境下更高速的扫描。

3.2.2 网络流量分析技术

被动探测方法主要采用的是网络流量分析技术。网络流量分析是指通过采集网络流量来进行一定的预处理，并采用分析算法对 7 层网络结构中各层的流量分布进行监测，对协议类型、流量内容进行综合分析的过程[36]。该技术通常用于运营商了解网络流量的分布和带宽使用情况，进行日常的运营和维护，也可用于网络管理员对网络的规划升级和性能优化。同时，在网络安全管理中，可用于加强管理员对网络行为的理解，检测隐藏的安全问题(服务异常使用、网络入侵等)。

基于网络流量分析技术对各层网络数据进行详细剖析的优势，可将其用于网络资产的被动探测，即以获取的网络流量为基础，进行主机、端口、操作系统、应用等网络资产的识别。本节首先分析了各层网络协议指纹所反映的资产信息，然后结合网络资产被动探测的实现，介绍了流量管理和入侵检测两类网络流量分析工具。

从 TCP/IP 体系使用的 4 层结构看，网络接口层(包括物理层和数据链路层)指纹主要体现了硬件设备间的差别，可用于识别网络资产类别，但只限于在本地局域网内获取和分析；网络层和传输层的数据可以更加便捷地在广域网络上传播，其参数一般用于操作系统的类型和版本的识别；在不考虑目标网络数据隐私保护的前提下，基于深度包检测技术(DPI)，结合软件版本与操作系统版本的映射关系，对应用层及以上协议中的特征字段或负载数据进行深度分析，可以实现软件版本的检测[37]。

根据流量采集深度的不同，用于被动资产发现的网络流量分析工具通常可分为流量管理类和入侵检测类。NetFlow 是流量管理类的典型代表，是 Cisco 为其设备(交换机和路由器)开发的一款网络流量管理工具[38]。其由一个流量出口、收集器及二者之间的专线组成，能够有效帮助管理员进行流量镜像、审计等工作。Klepsland 提出了一种基于 NetFlow 的被动资产探测方法[39]，通过对采集的流量特征信息加载不同的过滤规则可以实现资产探测。但由于该工具是为流量审计设计的，仅保留了网络流量的一些基础信息，因此其难以进行深入分析，且对于一些非标准端口背后的服务及其版本信息更是无能为力。

当前大部分的入侵检测系统都是以网络流量分析为基础的，基于特征的入侵检测系统 Snort、Suricata 适用于从网络流量中发现一些字节信息；而 Bro 通常用于处理更复杂的(需要更高水平协议知识的、贯穿多种网络流的)任务[2]。Bro 是一套事件引擎和策略分离的入侵检测系统，主要由基于 libpcap 的被动流量捕获部分、事件引擎、策略脚本 3 部分组成[40]。其支持大部分常见网络协议的识别，同时借助动态协议检测(Dynamic Protocol Detection)方法，对出现在非标准端口的网络流量依然可以进行一定程度的识别，支持的应用层协议和隧道协议有 DHCP，DNS，FTP，HTTP，SSH 等。Bro 将协议分析与内容分析分离，提高了处理效率，支持用户编写需要的过滤策略脚本，因此，将原本的入侵检测特征修改为需要的网络资产特征，便可实现资产探测的功能。Philip[37] 以软件生命周期为主线，对不同阶段网络资产通信流量特征进行分析，基于 Bro 实现了内网网络资产的被动探测。

3.2.3 网络资产指纹识别技术

在主、被动探测结果的基础上，要得到网络资产识别结果，离不开指纹特征匹配的过程，且识别结果的准确性很大程度上取决于特征匹配的准确性。由于网络技术在不同行业、不同领域(如工控网、物联网等)的广泛应用，网络服务日趋多样化，对每一类网络资产进行识别都需要提取专用、特殊的指纹特征信息，难以进行全面、准确的概括，但识别的方法具有通用性。因此，本节主要以最具代表性的操作系统指纹识别为例进行阐述。常见的指纹识别方法主要有常规特征匹配和基于机器学习算法的特征匹配两种。

1)常规指纹特征匹配

常用的操作系统指纹识别特征有：IP 头中的总长度(total length)、标志(ID)、是否分片(DF)、生存时间(TTL)字段等，TCP 头中的可选项，TCP 头部的窗口大小(wsize)、可选项等，以及 ICMP、UDP 协议指纹、SYN-ACK 包重传时延等。

在匹配方法上，除了最基础的精确匹配外，考虑到网络延时、动态配置等现实因素，类似正则表达式的模糊匹配方法也被广泛应用，如目前最新版的 p0fv3[17] 指纹库将操作系统指纹分为两类(specified，generic)，增加了用于模糊匹配的 g 类，在 s 类无法精确匹配的情况下，为避免直接给出未知的结果，对 mss，wsize，scale 等字段均允许使用“*”进行模糊匹配，进而粗略地给出目标操作系统所属的大类；SinFP[41] 中的启发式匹配机制也是一种分层次的模糊匹配。

此外，还有采用了综合加权匹配的方法，如 Nmap 的指纹评分表对各响应数据项的指纹识别能力进行了量化评分，在计算响应同该条指纹的匹配率时，引入各数据项的评分作为加权，从而进一步提高识别的准确率。

2)基于机器学习的特征匹配

随着数据挖掘技术的发展，国内外很多研究人员将机器学习方法引入到操作系统指纹识别领域，提高了指纹匹配的能力。Beverly 等[42] 引入朴素贝叶斯分类器进行操作系统指纹识别，实现了对未精确匹配指纹的识别，但其结果的准确性受限于样本分布的情况，具有一定的不确定性。Sarraute 等[43] 利用 Nmap 指纹库训练的神经网络模型识别操作系统指纹，比 Nmap 自身识别的准确率要高，但由于需要使用 RPC 服务识别版本，其适用范围受到限制。

Al-Shehari 等[44] 提出了一种基于 C4.5 决策树算法分类器的操作系统指纹识别方法，该方法利用 TCP 连接 socket 的 hash 值关联 SYN 包和 FIN 包，对 p0f 指纹库中的 SYN 包指纹进行扩展，加入了 FIN 包中的部分特征。扩展 p0f 指纹对识别的准确率略有提高，但指纹特征个数增加也意味着计算量的增加，在新版本 p0fv3 指纹库[17] 中已将 RST 和 FIN 特征删除，仅剩下 SYN 和 SYNACK 指纹，分别用于 TCP 连接的客户端和服务器的识别。

Tyagi 等[45] 提出了一种基于欧氏距离的操作系统识别方

法，用于发现企业内网的非授权主机，相比于其他复杂的分类器缩短了建模时间，但该方法使用的指纹特征数量有限，每种操作系统指纹只用了一条特征来描述，且仅用取对数的方法进行粗略的属性归一化，导致较小值域的属性（Nop，DF，Timestamp 等二元属性）权重过小，难以全面地区分不同类别的操作系统，识别的操作系统类型和版本信息不够全面，仅能做到粗粒度的识别。

国内的邹铁铮等[46]将 Nmap 指纹库样本向量化，利用支持向量机（SVM）方法实现了粗粒度的操作系统分类器，有效处理了未精确匹配指纹的识别问题，但输入向量维度过高，需要进一步处理。程书宝等[47]采用了奇异值分解和有向无环图的方法，先对初始操作系统指纹生成的矩阵进行奇异值分解，并提取奇异值特征，然后基于有向无环图生成的多类分类器对未知指纹的奇异值特征进行分类。该方法在降低向量维数的同时对未精确匹配的指纹也具有较高的识别率。易运晖等[48]利用 C4.5 决策树模型，以较短的建模时间和较高的准确率实现了基于 TCP/IP 协议栈指纹的被动操作系统识别，提高了未精确匹配指纹的识别率。

4　未来研究方向

基于上述对传统、新型两大类网络资产探测及其关键技术的分析，本文给出了下一步可能的研究方向。

4.1　多种类型探测方法的综合运用

现有不同类型的网络资产探测技术在具体的应用场景下各有所长，主动探测方法的探测速度快，且对不产生网络流量的资产也能有效探测，但噪声大，仅能了解探测当时的情况；被动探测虽噪声小，支持历史数据积累，但对不产生网络流量的资产无效；基于搜索引擎的非入侵式探测隐蔽、高效，但探测目标仅限于具有公网 IP 的资产，易受欺骗，准确度相比前两种方法较低。因此，如何进行综合运用，取长补短，以得到最佳的探测结果是值得研究的问题，相关研究人员也进行了一定的探索。

Auffret 提出基于 Perl 语言的 SinFP[41]，对被动探测方法分析的 TCP SYN 数据包指纹进行一定的修改，将其作为一种响应数据包融入到主动方式指纹中，采用启发式算法进行模糊匹配，从而兼顾了主、被动探测的优点，较好地处理了网络地址转换（NAT）和端口转换（PAT）的广泛应用给网络资产探测带来的问题。但该研究仅针对基于响应协议栈指纹的操作系统识别，应用范围有限。主动指纹通常是响应数据包类型、字段、时延等特征，而被动指纹仅限于部分协议字段特征，二者如何结合、结合后匹配率降低等问题仍需要深入研究。

Simon 等[49]综合运用谷歌和 Shodan 获取的目标域名信息，实现了对该域名的非入侵式资产探测和漏洞分析。在网络资产探测中，如何综合运用通用和网络安全专用搜索引擎也是值得研究的问题。但如 3.1.3 节所述，当前国内外相关的网络安全专用搜索引擎种类繁多，数据获取能力难以准确评估，如何根据实际应用场景的需求，合理选择某种或多种搜索引擎作为数据来源就显得异常重要。

Genge 等[50]将 p0f，PRADS 的被动探测结果与从 Shodan 获取的公网探测结果融合，对 5 所科研院校网络中的旧版本（存在安全威胁）的敏感服务进行了探测。联合基于搜索引擎的非入侵式方法和本地主、被动探测方法，可以实现从内、外网两个视角了解本地网络资产情况，提高了全面性，这也是一个重要的研究方向。

4.2　对具体应用场景和对象的探测

当前，互联网应用领域广泛拓展，网络资产的类型也日趋多样化。本文介绍的探测方法是相对基础的，在对不同类型的网络资产进行探测识别的过程中，具体的网络扫描、流量分析的协议内容等还需要根据实际的探测场景和对象进行有针对性的研究和开发。如对工控设备进行探测时，需要考虑其特有的通信协议（Modbus，DNP3，Siemens S7 等），选取出有效的特征，再基于本文所阐述的基础方法实现探测。

4.3　人工智能在指纹识别中的应用

在网络资产指纹识别方面，3.2.3 节介绍了基于机器学习方法的网络资产指纹匹配的相关研究[42-48]，这些研究大多是直接引用现有探测工具指纹库中使用的特征，未结合实际探测需求对这些特征进行取舍和补充；采用的算法也大多是机器学习中的基础算法，准确率有待提高，但带来的运算开销却不小，难以实际应用；此外，目前没有可供直接使用的网络资产探测相关数据集，而依靠人工标记的方法工作量又很大，现有研究通常以指纹库中的指纹为训练集，识别的精度受到限制。因此，如何结合网络资产指纹识别的需求，更好地借助人工智能技术提高识别的精度和效率，还有很多值得思考和探索的问题。

结束语　本文回顾了网络资产探测技术的发展历程和研究现状，介绍了传统和新型两类网络资产探测方法，其中新型类方法包括主动、被动和基于搜索引擎的非入侵式探测，并归纳总结了各种方法的优缺点；同时，分析了涉及的高速网络扫描、网络流量分析、网络资产指纹识别等关键技术。多种探测方法的综合运用、对具体场景和对象的探测、人工智能在指纹识别中的应用等都是未来可能的研究方向。作为开展网络安全管理的重要基础，网络资产探测技术将得到持续的研究和关注。

参 考 文 献

[1] International Organization for Standardization. ISO/IEC 13335-1:2004 [EB/OL]. https://www.iso.org/standard/39066.html.

[2] SANDERS C, SMITH J. Applied Network Security Monitoring: Collection, Detection, and Analysis[M]. Syngress Publishing, 2013:3-5.

[3] HAUKELI J. False positive reduction through IDS network awareness[D]. Oslo: University of OSLO, 2012.

[4] YE Z Y, GUO Y B, WANG C D, et al. Survey on application of attack graph technology[J]. Journal on Communications, 2017, 38(11):121-132. (in Chinese)
叶子维，郭渊博，王宸东，等. 攻击图应用研究综述[J]. 通信学报，2017，38(11)：121-132.

[5] 吴灏. 网络攻防技术[M]. 北京：机械工业出版社，2009：10-14.

[6] SCOTT A, JAY H, GREG K, et al. Spiceworks homepage[EB/OL]. https://www.spiceworks.com.

[7] BORANBAYEV A S. Defining methodologies for developing

J2EE web-based information systems[J]. Nonlinear Analysis Theory Methods & Applications,2009,71(12):e1633-e1637.

[8] LAUFER K. A Hike through Post-EJB J2EE Web Application Architecture[J]. Computing in Science & Engineering,2005,7(5):80-88.

[9] LYON G F. Nmap Network Scanning:The Official Nmap Project Guide to Network Discovery and Security Scanning[M]. Insecure,2009.

[10] YAROCHKIN F V,ARKIN O,KYDYRALIEV M,et al. Xprobe2++:Low volume remote network information gathering tool[C]// IEEE/IFIP International Conference on Dependable Systems & Networks. IEEE,2009:205-210.

[11] SHAMSI Z,NANDWANI A,LEONARD D,et al. Hershel:Single-Packet OS Fingerprinting[J]. IEEE/ACM Transactions on Networking,2016,24(4):2196-2209.

[12] SHAMSI Z,LOGUINOV D. Unsupervised Clustering Under Temporal Feature Volatility in Network Stack Fingerprinting [J]. IEEE/ACM Transactions on Networking,2016,PP(99):1-14.

[13] GARCIA S. DNmap:the distributed nmap[EB/OL]. http://mateslab. weebly. com/dnmap-the-distributed-nmap. html.

[14] DURUMERIC Z,WUSTROW E,HALDERMAN J A. ZMap:fast internet-wide scanning and its security applications[C]// Usenix Conference on Security. San Jose:USENIX Association,2013:605-620.

[15] GRAHAM R D. Masscan:the entire Internet in 3 minutes [EB/OL]. http://blog. errat asec. com/2013/09/masscanentire-internet-in-3-minutes. html?utm_source=tuicool&utm_medium=referral#. V9AqVLG8rzI.

[16] ADRIAN D,DURUMERIC Z,SINGH G,et al. Zippier ZMap:internet-wide scanning at 10 Gbps[C]// 8th USENIX Workshop on Offensive Technologies (WOOT 14). 2014.

[17] ZALEWSKI M. p0f v3:Passive fingerprinter [EB/ OL]. http://lcamtuf. coredump. cx/p0f3.

[18] BARNES J,CROWLEY P. k-p0f:a high-throughput kernel passive os fingerprinter[C]// Proceedings of the Ninth ACM/IEEE Symposium on Architectures for Networking and Communications Systems. IEEE Press,2013:113-114.

[19] CHEN J,WAN Y P,CHEN H,et al. Research on High-Performance Operating System Detection Method[J]. Journal of University of South China (Science and Technology),2016,30(1):66-70. (in Chinese)
陈军,万亚平,陈虹,等. 高性能操作系统检测方法研究[J]. 南华大学学报(自然科学版),2016,30(1):66-70.

[20] FJELLSKAL E. Passive real-time asset detection system[EB/OL]. http://gamelinux. github. io/pr ads.

[21] SHELTON M. Passive asset detection system[EB/OL]. http://passive. sourceforge. net/about. php.

[22] FALCH P B. Investigating passive operating system detection [D]. University of OSLO Department of Informatics,2011.

[23] KOLLMANN E. Chatter on the Wire:How Excessive Network Traffic Gives Away Too Much! [EB/OL]. http://chatteronthewire org.

[24] HJELMVIK. Networkminer homepage [EB/OL]. http://networkminer. sourceforge. net.

[25] WANG C D,GUO Y B,HUANG W. Non-intrusive Network Security Scanning Technology[J]. Information Security and Communications Privacy,2016(9):67-72. (in Chinese)
王宸东,郭渊博,黄伟. 非入侵式网络安全扫描技术研究[J]. 信息安全与通信保密,2016(9):67-72.

[26] HUANG C. Research and Practice of Vulnerability Scanning Technology Based on GHDB [D]. Beijing:Beijing Jiaotong University,2012. (in Chinese)
黄超. 基于 GHDB 的漏洞扫描技术的研究与实践[D]. 北京:北京交通大学,2012.

[27] MATHERLY J. Shodan tool[EB/OL]. https://www. shodan. io.

[28] DURUMERIC Z,ADRIAN D,MIRIAN A,et al. A Search Engine Backed by Internet-Wide Scanning[C]// ACM Sigsac Conference on Computer and Communications Security. Colorado:ACM,2015:542-553.

[29] 404 Team from Knownsec. ZoomEye search engine[EB/OL]. https://www. zoomeye. org.

[30] DUGGAN D P. Penetration Testing of Industrial Control Systems[R]. Sandia National Lab,2005:5-7.

[31] GENGE B,GRAUR F,ENĂCHESCU C. Non-intrusive Techniques for Vulnerability Assessment of Services in Distributed Systems[J]. Procedia Technology,2015,19:12-19.

[32] MATHERLY J. Complete Guide to Shodan[OL]. http://leanpub. com.

[33] LAB B. Report on the Organizational Behavior of Key Infrastructure Information Collection in Cyberspace[EB/OL]. (2016-05-03) [2017-09-26]. http://plcscan. org/blog/wpcontent/uploads/2016/06/ics-security-research-report-2016-05. pdf.

[34] 李瑞民. 网络扫描技术揭秘[M]. 北京:机械工业出版社,2012:1-18.

[35] MYERS D,FOO E,RADKE K. Internet-wide scanning taxonomy and framework[C]// Proceedings of Australasian Information Security Conference (ACSW-AISC). Australian Computer Society,Inc,2015.

[36] 周涛. 网络安全中的数据挖掘技术[M]. 北京:清华大学出版社,2017:162-167.

[37] PHILIP C S. IDS-based Passive Asset Detection:Using and extending an IDS for asset detection [D]. University of OSLO Department of Informatics,2014.

[38] Cisco. Introduction to cisco ios netflow [EB/OL]. http://www. cisco. com/c/en/us/products/ios-nx-os-software/ios-netflow/index. html.

[39] KLEPSLAND M E. Passive Asset Detection using NetFlow [D]. University of OSLO Department of Informatics,2012.

[40] PAXSON V. Bro:a system for detecting network intruders in real-time[J]. Computer Networks,1999,31(23-24):2435-2463.

[41] AUFFRET P,SINF P. Unification of active and passive operating system fingerprinting[J]. Journal of Computer Virology and Hacking Techniques,2010,6(3):197-205.

[42] BEVERLY R. A Robust Classifier for Passive TCP/IP Fingerprinting[C]// Passive and Active Network Measurement,International Workshop. DBLP,2004:158-167.

[43] SARRAUTE C, BURRONI J. Using Neural Networks to improve classical Operating System Fingerprinting techniques[J]. Computer Science, 2008, 8(1): 35-47.

[44] AL-SHEHARI T, SHAHZAD F. Improving Operating System Fingerprinting using Machine Learning Techniques[J]. International Journal of Computer Theory & Engineering, 2014, 6(1): 57-62.

[45] TYAGI R, PAUL T, MANOJ B S, et al. Packet Inspection for Unauthorized OS Detection in Enterprises[J]. IEEE Security & Privacy Magazine, 2015, 13(4): 60-65.

[46] ZOU T Z, LI Y, ZHANG B F, et al. Operating system recognition based on support vector machines [J]. Journal of Tsinghua University (Natural Science Edition), 2009(s2): 2164-2168. (in Chinese)
邹铁铮，李渊，张博锋，等. 基于支持向量机的操作系统识别方法[J]. 清华大学学报(自然科学版)，2009(s2)：2164-2168.

[47] CHEN S B, HU Y. Operating System Recognition based on Singular Value Decomposition and DAG_SVMS[J]. Information Security and Communications Privacy, 2013(9): 66-67. (in Chinese)
程书宝，胡勇. 基于奇异值分和 DAG_SVMS 的操作系统类型识别[J]. 信息安全与通信保密，2013(9)：66-67.

[48] YI Y H, LIU H F, ZHU Z X. Research of Passive OS Recognition Based on Decision Tree [J]. Computer Science, 2016, 43(8): 79-83. (in Chinese)
易运晖，刘海峰，朱振显. 基于决策树的被动操作系统识别技术研究[J]. 计算机科学，2016，43(8)：79-83.

[49] SIMON K, MOUCHA C, KELLER J. Contactless Vulnerability Analysis using Google and Shodan [J]. Journal of Universal Computer Science, 2017, 23(4): 404-430.

[50] GENGE B, HALLER P, ENĂCHESCU C. Beyond Internet Scanning: Banner Processing for Passive Software Vulnerability Assessment[J]. International Journal of Information Security Science, 2015, 4(3): 81-91.

(上接第 23 页)

[21] YUSUKE I, TOSHIAKI M. A DTN routing algorithm adopting the "Community" and "Centrality" parameters used in social networks[C]// 2018 International Conference on Information Networking (ICOIN). New York: IEEE, 2018: 211-216.

[22] ALAOUI E A A, AGOUJIL S, HAJAR M, et al. Improving the data delivery using DTN routing hierarchical topology (DRHT) [C]// International Conference on Wireless Networks and Mobile Communications. New York: IEEE, 2016: 1-5.

[23] ALAOUI S E, RAMAMURTHY B. Routing optimization for DTN-based space networks using a temporal graph model[C]// IEEE International Conference on Communications. New York: IEEE, 2016: 1-6.

[24] MTIBA A, MAYM, DIOTC, et al. People rank: social opportunistic forwarding [C]// Conference on Information Communications. New York: IEEE Press, 2010: 111-115.

[25] WANG G Z, YAN L J, ZHENG L, et al. Social routing based on location preference prediction in DTN[C]// 2017 11th IEEE International Conference on Anti-counterfeiting, Security and Identification(ASID). New York: IEEE, 2017: 154-157.

[26] LIU C, WU J. Routing in a cyclic mob space [C]// ACM International Symposium on Mobile Ad Hoc NETWORKING and Computing. New York: ACM, 2008: 351-360.

[27] QI W, SONG Q, WANG X, et al. Trajectory Data Mining-Based Routing in DTN-enabled Vehicular Ad Hoc Networks[J]. IEEE Access, 2017, PP(99): 128-138.

[28] LIU J, TANG M, YU G. Adaptive Spray and Wait Routing Based on Relay-Probability of Node in DTN [C]// 2012 International Conference on Computer Science & Service System (CSSS). New York: IEEE, 2012: 1138-1141.

[29] WANG G Z, ZHENG L, YAN L J, et al. Probabilistic routing algorithm based on transmission capability of nodes in DTN[C]// 2017 11th IEEE International Conference on Anti-counterfeiting, Security and Identification (ASID). New York: IEEE, 2017: 146-149.

[30] HUANG Z, ZHANG Q, XIN X, et al. DTN routing algorithm based on service probability and limited copy for satellite networks[C]// International Conference on Optical Communications and Networks. New York: IEEE, 2017: 1-3.

[31] CONG L, YANG H, WANG Y, et al. Research on the Routing Algorithm of LEO-Satellite DTN Network Based on Multi-Attribute Decision Making[C]// International Conference on Intelligent Computation Technology and Automation. New York: IEEE, 2017: 166-171.

[32] HE J, XU C, WU Y. Resource-efficient routing protocol based on historical encounter time interval in DTN[C]// IEEE International Conference on Computer and Communications. New York: IEEE, 2017: 2026-2031.

[33] LIN F, WANG Y, WU H. Testbed Implementation of Delay/Fault-Tolerant Mobile Sensor Network (DFT-MSN) [C]// IEEE International Conference on Pervasive Computing and Communications Workshops. New York: IEEE, 2006.

[34] ZHAO W. A message ferrying approach for data delivery in sparse mobile ad hoc networks [C]// ACM International Symposium on Mobile Ad Hoc Networking and Computing. New York: ACM, 2004: 187-198.

[35] ZHAO W, AMMAR M, ZEGURA E. Controlling the mobility of multiple data transport ferries in a delay-tolerant network [C]// INFOCOM 2005. Joint Conference of the IEEE Computer and Communications Societies. New York: IEEE, 2005: 1407-1418.

[36] ZHANG H, LU N, MA J, et al. HNSARA: A history-and-node-state-based active routing algorithm for DTN[C]// International Conference on Cyberspace Technology. England: IET, 2015: 1-5.

[37] KAWAMOTO Y, NISHIYAMA H, KATO N. Toward terminal-to-terminal communication networks: A hybrid MANET and DTN approach[C]// IEEE, International Workshop on Computer Aided Modeling and Design of Communication Links and Networks. New York: IEEE, 2014: 228-232.