网络资产探测关键技术研究*

邵 磊[1],余 晓[2],吴剑章[1]
(1.东南大学 网络空间安全学院,江苏 南京 210096;2.东南大学 继续教育学院,江苏 南京 210096)

摘 要: 网络资产探测是指探测具有网络连接的硬件和软件,可以为企业、学校等网络资产全生命周期管理打下坚实的基础。结合资产标识、漏洞管理能够有效追溯每个资产的状态并及时弥补漏洞,避免造成经济损失。从网络资产探测相关技术角度出发,将相关技术分类为扫描识别技术、性能优化技术、隐蔽扫描技术和机器学习技术,先阐述当前技术面临的挑战,然后分析并总结各类技术的原理及其优缺点,最后对未来研究方向进行展望。

关键词: 网络资产探测;操作系统识别;流量分析;搜索引擎;机器学习
中图分类号: TP393.0 **文献标识码:** A **DOI:** 10.19358/j.issn.2097-1788.2022.05.001
引用格式: 邵磊,余晓,吴剑章. 网络资产探测关键技术研究[J].网络安全与数据治理,2022,41(5):3-9,35.

Research on method of network assets detection

Shao Lei[1], Yu Xiao[2], Wu Jianzhang[1]
(1.School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China;
2.School of Continuing Education, Southeast University, Nanjing 210096, China)

Abstract: Network asset detection refers to the detection of hardware and software connected with network, which can lay a solid foundation for the full lifecycle management of network assets in enterprises, schools, etc. Combined with asset identification and vulnerability management, it can effectively trace the status of each asset and make up for vulnerabilities in time to avoid economic losses. From the perspective of technologies related to network asset detection, the technologies are classified as scanning technology, performance optimization technology, stealthy scanning technology and machine learning technology. Firstly, the challenges faced by current technologies are elaborated. Secondly, the principles, advantages and disadvantages of these technologies are analyzed. Finally, the future research directions are summarized.

Key words: network asset detection; operating system identification; traffic analysis; search engine; machine learning

0 引言

网络资产探测的目的是探测目标范围内在线的主机及其开放的端口、服务、操作系统等信息。掌握资产信息是进行网络资产全生命周期管理的前提,结合资产标识技术可以追踪任何一个网络资产的在线情况、版本信息,为更新版本、下线不再使用的资产等管理工作提供信息基础。还可以结合最新的漏洞信息,更好地保护面临威胁的网络资产,避免被不法分子利用而造成财产损失。

1 网络资产探测概述

网络资产可以定义为具有网络连接的任何对于企业、高校等组织具有价值的资源。网络资产主要可以分类为硬件和软件,硬件包含各种通信、计算、存储类等设备;软件主要是指运行在硬件上的各类服务,例如 Web 服务、数据库、操作系统等。

网络资产探测方法可以分类为:主动探测、被动探测和基于安全搜索引擎探测[1]。本文根据各类方法所需要使用的相关技术,从技术角度分类为扫描识别技术、性能优化技术、隐蔽扫描技术和机器学习技术。扫描识别指的是直接进行资产探测的相关技术,如通过主动发送数据包进行扫描。性能优化是指对于扫描技术在效率上的优化技术。由于资产探测往往伴随着被防火墙、入侵检测系统等发现的风险,如对一段互联网协议(Internet Protocol, IP)

* 基金项目:中国高校产学研创新基金(2020ITA07007)

地址进行顺序扫描很容易被网管察觉并警告,这就需要使用到隐蔽扫描。现如今,机器学习的应用范围不断扩大,而在网络资产方面,它可以应用于根据常见协议的指纹特征进行操作系统识别。

2 相关技术

2.1 扫描识别技术

2.1.1 网络协议

常见可用于网络探测的协议有:地址解析协议(Address Resolution Protocol,ARP)、网际控制报文协议(Internet Control Message Protocol,ICMP)、传输控制协议(Transmission Control Protocol,TCP)、用户数据报协议(User Datagram Protocol,UDP)、简单网络管理协议(Simple Network Management Protocol,SNMP)等。

ARP作为网络层的协议,其功能是将网络层使用的IP地址解析为数据链路层使用的介质访问控制(Media Access Control,MAC)地址。给目标主机发送ARP请求,如果目标主机响应请求返回了MAC地址,说明目标主机是在线的,否则说明目标主机不在线,所以ARP协议可以用于主机扫描,Arping命令就是通过该协议实现的。它的优点在于局域网内精度很高,且可以突破防火墙[2],然而应用范围较窄,只能在局域网内使用。

ICMP属于网络层的协议,差错报告报文和询问报文是该协议的两个不同种类。ICMP的差错报告报文和询问报文都可以用来进行主机扫描,只要目标主机有响应就说明该主机在线。Jiang Weihua[3]等人就使用ICMP报文进行了主机扫描,甚至通过分析ICMP数据包的IP首部特点来辨别不同的操作系统[3-4]。Ping和Traceroute也是该协议的应用,但是由于Ping工具使用率很高,许多注重安全的主机防火墙都不允许通过ICMP报文。

TCP是传输层的协议,它提供了可靠的面向连接的服务。通过发送TCP报文到目标主机的目标端口,既可以确定该主机是否在线,也可以确定目标端口是否打开。TCP扫描主要有:Connect扫描、SYN扫描、ACK扫描、FIN扫描、NULL扫描、XMAS扫描等。

TCP Connect扫描又叫全连接扫描,顾名思义,就是源主机和目标主机进行三次握手建立完整的连接,根据目标主机的不同反应,结果可能分为三种情况:(1)建立了完整连接,既确定了主机在线,也确定了目标端口是开放的;(2)源主机发出SYN请求后,目标主机返回了带RST标记的响应报文,虽然表明了目标端口是关闭的,但是可以确认目标主机在线;(3)源主机发出SYN请求后,目标主机没有回应,可能该主机不在线,也可能目标端口受到了防火墙保护。TCP Connect扫描实现简单,但是容易在目标主机中留下大量TCP建立连接的日志,易被发现。

TCP SYN扫描,又称为半连接扫描,源主机和目标主机并不会建立一个完整的TCP连接。源主机向目标主机发送SYN请求后,如果目标主机返回SYN/ACK响应,表明目标端口处于打开状态,之后源主机发送RST报文断开连接。如果目标主机响应的是RST报文,说明目标端口处于关闭状态,如果没有响应,说明目标主机不在线或被防火墙保护。SYN扫描的优点在于不会在目标主机留下相关日志,相比全连接扫描更加隐蔽。

TCP ACK扫描是源主机发送一个只设置ACK标志位的请求,不管目标主机的端口是否打开,该主机都会响应一个RST数据包,而如果没有收到响应,说明目标主机可能不在线或者受到防火墙保护,如果收到ICMP终点不可达的差错报文,说明目标主机受到了防火墙保护。考虑到目标主机如果响应了RST报文,则说明该主机一定在线,所以ACK扫描可以用于主机扫描。

TCP FIN扫描是源主机发送一个带有FIN标记的数据包到目标主机端口,通常情况下,如果目标端口是关闭的,那么目标主机会发送一个RST响应包。如果主机没有任何反应,说明端口打开或者受防火墙保护。如果收到ICMP终点不可达的差错报文,说明目标受到防火墙保护。TCP FIN扫描的优点是更加隐蔽,不会留下痕迹,但是容易受到误导,因为在收不到响应时无法区分目标端口是打开的还是受到了防火墙保护,而且有的操作系统在端口打开的时候也会发送RST响应。SYN扫描如果和FIN扫描结合起来使用可以简单地做一些操作系统识别,但是识别能力有限。

TCP NULL扫描是将发送的TCP数据包首部标记都置为0,TCP XMAS扫描则是将FIN、URG、PSH标记都置为1,它们与FIN扫描类似,不同操作系统可能有不同的实现。通常情况下,如果目标主机端口处于关闭状态,会发送一个RST响应包,如果没有响应说明端口打开或者受到防火墙保护,如果收到了ICMP终点不可达的差错报文,说明端口受到

防火墙保护。

一些常见的扫描工具如 Nmap [5]、Zmap[6]、Masscan[7]都使用了 TCP 协议进行主机、端口扫描。Nmap 功能强大,它支持以上所有类型的扫描,优点是搜索结果较为全面,缺点是速度较慢。DNmap[8]使用分布式架构,对 Nmap 的探测速度进行了改进。Yuan[9]等人使用 Go 语言结合 TCP 全连接扫描开发了名为 HIRFL 的扫描器,其扫描速度比 Nmap 快。Beverly[10]基于 TCP SYN 和 ACK 扫描开发了 Yarrp,能够以 100 kpps 的速度在 30 min 内发现超过 40 万个路由器接口。Li[11-12]在 TCP 扫描的基础上,使用了可编程的交换机开发了快速可扩展的 IMap 扫描器。Zhang 等通过将 Zmap 和 Nmap 相结合[13],开发了扫描 VxWorks 主机信息的软件,并提供人性化的管理界面和扫描报告。Sivanathan[14]等通过 TCP 全连接的方式将 TCP 扫描用于物联网设备的识别和分类。

UDP 是无连接、尽最大努力交付的传输层协议。UDP 扫描一般是利用 ICMP 协议进行的,源主机向目标主机端口发送一个长度为 0 的 UDP 数据包,如果收到了 ICMP 终点不可达的差错报告报文,说明目标端口是关闭的。如果收到了 UDP 响应报文,说明目标端口是打开的。如果没有响应,那么目标端口可能打开着,也可能受到了防火墙保护。Nmap 也具备 UDP 扫描功能,使用“-sU”参数便可以进行 UDP 扫描。Jung[15]等基于 UDP 协议提出了一种新的技术用于扫描物联网设备,它包含了主扫描和辅助扫描两个部分,主扫描使用简单服务发现协议(Simple Service Discovery Protocol,SSDP),辅助扫描使用名称服务器协议(Network Basic Input/Output System Name Service,NBNS)和多播域名解析协议(Multicast Domain Name System,MNDS),平均速度比 Nmap 快 1 524 倍。Kumar[16]等基于 UDP 协议设计了使用多个 IP 地址扫描目标端口的扫描器,比传统扫描方法快约 190 倍。

SNMP 是基于 UDP 协议设计的用于管理网络节点的应用层协议,可以实现对网络中的资源进行监控。现在已经有了三个版本:SNMPv1、SNMPv2c、SNMPv3。SNMP 三个主要组件分别是:管理信息结构(Structure of Management Information,SMI)、管理信息库(Management Information Base,MIB)和管理协议,其中管理信息结构和管理信息库是 ASN.1 语法定义的模块。SNMP 用对象标识符(Object Identifier,OID)唯一标识一个被管理对象,所有被管理对象的集合就是管理信息库,它是一个树型结构,所以又称 MIB 树。管理信息结构的工作是为管理信息库定义被管理对象的标准、规定数据类型以及分配 OID。网络管理者、代理者、被管理设备共同组成了 SNMP 所管理的网络,代理者和管理者通过协议数据单元(Protocol Data Unit,PDU)进行通信。在进行网络探测时,MIB 库中的 ipAddrTable、ipRouteTable、ipNetToMediaTable 是需要关注的三个对象,ipAddrTable 对象中包含了每个接口对应的 IP 地址,ipRouteTable 对象可以获取到直接与该设备连接的子网和路由器信息,ipNetToMediaTable 记录了该设备路由表中的 ARP 映射。Handoko[17]等通过使用 SNMP 创建实时网络拓扑,设计并开发了可以监控网络节点信息的应用程序。Saheb[18]等通过结合 ICMP 和 SNMP 协议实现了网络资源的自动发现和监控。Roquero[19]等提出了一种具有灵活的多线程体系结构的 SNMP 管理器体系结构,并使用自适应轮询算法有效减少了轮询不断增加的 SNMP AGENT 带来的资源消耗。使用 SNMP 可获得足够多的主机信息用于资产管理,但是现有路由器大多默认不打开 SNMP 协议,且需要知道 Community 的值才能访问 MIB。

2.1.2 流量分析

流量分析往往用于被动探测,通过在网络节点部署探针来捕获数据,分析这些网络流量的特征可以确定流量来源主机的某些特征[20]。不同操作系统对于 TCP/IP 协议栈的实现有所不同,可以根据 TCP/IP 数据报的指纹特征来区分操作系统的类别,常用的指纹特征有窗口大小、是否分片、数据包生存时间、最大报文长度、可选字段等。

应用层服务一般会使用较为固定的端口,可以通过 TCP/IP 数据包中的端口号来进行服务识别,其结果一般是较为准确的,服务及其对应的端口号见表 1。

常见的通过分析流量进行操作系统识别的工具有:Nmap、P0f[21]、Ettercap[22]、Satori[23]、NetworkMiner[24]等。Nmap 为了获得所需要的 TCP/IP 指纹信息,会发送多达 16 个请求数据包,其中包括 13 个 TCP 请求、2 个 ICMP 请求和 1 个 UDP 请求。P0f 从被动收集的 TCP 数据包分析流量来源主机可能的操作系统类型,不过准确度不如 Nmap。Bai[25]等基于 P0f 和可编程交换机,使用 P4 语言开发了P40f 工具,提供了数据包处理速率高达 100 Gpps 的情况下也能使

表 1 常见服务对应的端口号

服务	端口号
域名解析(DNS)	53
动态主机配置协议(DHCP)	67/68
简单网络管理协议(SNMP)	161/162
文件传送协议(FTP)	20/21
远程终端协议(TELNET)	23
超文本传送协议(HTTP)	80
简单邮件传送协议(SMTP)	25
邮件读取协议(POP3)	110
网际报文存取协议(IMAP)	143

用的操作系统识别功能。Ettercap 是一个开源项目,它可用于检测中间人攻击,也支持使用 TCP 指纹特征进行操作系统检测。Satori 通过分析 DHCP 协议的 Option 字段来识别操作系统。NetworkMiner 也是一个开源的操作系统识别工具,它使用了来自 P0f、Ettercap 和 Satori 的数据库,还利用了 Nmap 提供的 MAC 供应商列表。此外,Lastovicka[26]等提出了结合 TCP/IP 协议指纹、HTTP USER-AGENT 字段和与操作系统有关的特定域名进行操作系统识别的方法。

2.1.3 搜索引擎

搜索引擎,如百度、谷歌等是互联网上最常见的工具,不同于这类普通搜索引擎,网络安全搜索引擎如 Shodan,可以搜索到主机及其打开的端口、服务等,除此之外,还能搜索到暴露公网的硬件设备,如路由器、打印机、网络摄像头等。Shodan 支持关键词搜索,如 country:"CN",可以搜索到位于中国的网络资产。除了使用 Shodan 网页进行搜索,还可以注册并生成一个 API key 在编程开发时直接调用 Shodan 的搜索接口。Zolotykh[27-28]对 Shodan 的扫描行为进行了分析,发现其使用 TCP 和 UDP 协议进行扫描,且使用了许多强度不同的爬虫进行端口探测和 Banner 信息抓取。除了 Shodan 之外,常见的网络安全搜索引擎还有结合了 Zmap 的 Censys、结合了 Nmap 的 ZoomEye 等。搜索引擎对于暴露于公网的网络资产有着强大的搜索能力,且隐蔽性强,缺点在于还无法对内网资产进行有效探测。

2.2 性能优化技术

2.2.1 无状态技术

通常的 TCP SYN 扫描需要建立一个完整的会话,源主机向目标主机发送 SYN 请求后,如果目标主机返回 SYN/ACK 响应,表明目标端口处于打开状态,之后源主机发送 RST 报文断开连接。Zmap 和 Masscan 使用了 TCP SYN 无状态扫描,无状态指的是它们使用独立的发送和接收线程,在发送 SYN 数据包时,对探测地址进行哈希编码,而后发送 RST 取消连接,数据包发出后并不记录目标主机的回执状态,而是在目标响应 SYN/ACK 数据包时对其中的源 IP、源端口等信息进行校验,这样避免了存储连接状态的开销,缺点是会丢失大约 2%的数据[6]。

2.2.2 并行化技术

为了进一步提高扫描速度,研究人员引入了并行化技术,如针对传统扫描工具 Nmap 速度不够快的问题,开发了使用分布式架构的 DNmap[8]工具,提高了网络扫描效率。Zmap 的开发者在文献[29]中对随机 IP 生成操作进行分片,每组都能生成与其他组互不相交的随机 IP 子集,使得该工具可以应用到多线程中。胡栋梁[30]等使用了三层模型的分布式架构,利用消息中间件改进了传统单点主动扫描技术,降低了 CPU 使用率和扫描响应时间。

2.2.3 零拷贝技术

常用的探测工具使用 Libpcap 进行数据包的捕获与发送,但数据采集性能不佳。PF_RING[31]被提出来提高数据采集性能,它是一种新型的套接字,其对性能的提高主要靠轮询(Polling)、内存映射(Mmap)和环形缓冲区(Ring buffer)来实现。PF_RING 的结构如图 1 所示,在创建套接字时环形缓冲区也被创建出来,数据包在到达网卡后,处理器以轮询的方式将数据写入缓存区,使用轮询方式的优点是可以减少处理器响应中断的时间。内存映射使得应用程序可以直接读取内核态环形缓冲区的数据,避免了将

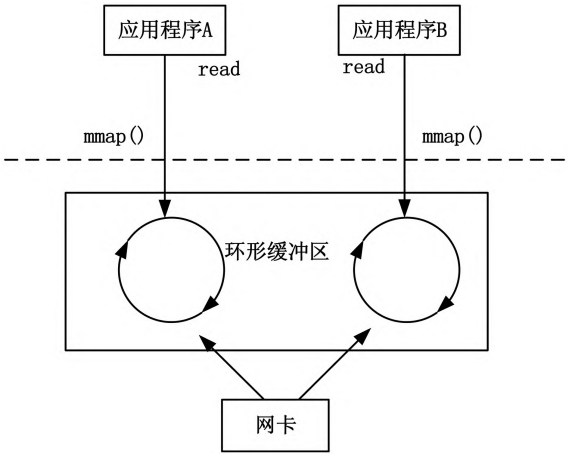

图 1 PF_RING 结构

数据包从内核态到用户态的一次拷贝。为了进一步减少开销，PF_RING ZC(Zero Copy)实现了直接网卡访问(Direct NIC Access，DNA)[32-33]，将网卡内存和寄存器直接映射到用户空间，支持了零拷贝技术，Zmap和Masscan都支持PF_RING ZC模式。面对网络流量突发而出现随机丢包的情况，Ye[34]等人在PF_RING的基础上设计了可以自动调节内核缓冲区空间大小的PF_RING-TA。

2.3 隐蔽扫描技术

2.3.1 随机IP地址生成

考虑到如果对一个IP地址空间的所有IP进行顺序扫描很容易被入侵检测系统或防火墙发现，所以有必要对所有IP打乱顺序，生成随机IP地址的方法主要有Zmap使用的基于素数原根的方法和Masscan使用的加密算法。Zmap的开发者在文献[29]中又进一步对该操作进行分片，每组都能生成与其他组互不相交的随机IP子集，使得该方法可以应用到多线程中。Beverly[10]开发的工具则通过加密算法生成随机的IP地址和生存时间(Time To Live，TTL)，使源主机能够以随机的TTL值和打乱的IP地址扫描目标地址空间。

2.3.2 多源IP扫描

为隐蔽扫描行为，可以使用多个源IP进行扫描，稀释单个IP被发现的风险，使得各IP看起来像在进行正常通信。据研究，安全搜索引擎Shodan使用了超过40个IP地址进行扫描，而这些IP地址分布在超过10个国家和地区[27-28]。

2.4 机器学习技术

对于操作系统的识别，除了使用常规的TCP/IP指纹库进行匹配之外，将机器学习应用于操作系统识别的研究也越来越多。Lastovicka[35]等人选取TTL、TCP窗口大小、初始SYN包大小作为特征，对比了朴素贝叶斯、决策树、K最近邻算法、支持向量机四种机器学习算法，得出了决策树最适合进行操作系统识别的结论。Fan[36]等人通过提取IP协议、TCP协议和TLS协议数据包的指纹特征，使用LightGBM算法(主要思想是利用弱分类器进行迭代以获得最优模型)，通过实验证明该算法的操作系统类型识别率达96.35%，操作系统版本识别率达84.72%，优于Lastovicka在文献[35]中得出的最优算法。Song[37]等人基于端口号、TTL、是否分片、窗口大小等特征，对比了传统的指纹匹配、K最近邻算法、决策树和神经网络算法对操作系统的识别率，发现神经网络算法的精确度达94%，高于其他识别方法。Kumar[38]等基于SYN包大小、TTL、窗口大小、HTTP USER-AGENT等特征，对比了K最邻近算法、人工神经网络、决策树、朴素贝叶斯、随机森林5种算法，其中K最邻近算法以96.22%的准确率成为操作系统识别最好的算法，并得出对分类结果影响最大的参数是TTL。可以看出，选取TCP/IP数据报首部的不同字段或者结合其他协议的某些字段都会影响机器学习在操作系统的识别率，如何选择合适的指纹特征和机器学习算法是今后的重要课题。

3 总结

网络资产探测相关的三类技术总结如下：

扫描识别相关的技术主要有网络协议、流量分析、搜索引擎，主要目标是扫描在线的主机、端口以及对运行的服务、操作系统进行识别。网络协议主要指的是通过ARP、ICMP、TCP等协议进行网络资产探测，优点是速度快、应用范围广、可用于不自己产生流量的资产，缺点在于可能被防火墙阻挡、被入侵检测系统发现、难以探测受到保护的资产。流量分析指的是通过收集网络流量，分析其源头的资产信息，优点是不容易被发觉、受保护的资产也有一定发现能力，缺点是主要用于内网、速度取决于探针的位置、发现不了不产生流量的资产。搜索引擎指的是使用Shodan等网络安全搜索引擎探测网络资产，优点是速度快、对于公网设备具有探测能力，缺点是不能用于内网、探测结果取决于搜索引擎。对比三种方法，从应用范围角度，网络协议众多，既可用于公网，也可用于内网，流量分析主要应用于内网，搜索引擎主要应用于公网；从速度角度，网络协议和搜索引擎速度较快，流量分析速度取决于探针位置；从隐蔽性角度，流量分析和搜索引擎不直接访问目标，隐蔽性高于主动发送数据包的网络协议。

性能优化技术包括无状态技术、并行化技术、零拷贝技术，主要目标是优化主机、端口扫描的速度。无状态技术是指不保存TCP SYN连接状态，使用独立的发送和接收线程，优点是避免了存储连接状态的开销，缺点是会丢失一些数据包。并行化技术是指将并行化引入网络资产探测，优点是可以提高探测速度，缺点是增加了成本。零拷贝技术是指避免网卡到用户态的数据拷贝，优点是减少了数据

包拷贝带来的开销,缺点是依然可能产生丢包。对比三种方法,从数据完整性角度,无状态和零拷贝技术都会产生丢包问题,并行化技术不会;从开销角度,无状态和零拷贝技术都减少了开销,而并行化技术增加了开销。

隐蔽扫描技术包括随机 IP 地址生成和多源 IP 扫描,主要目标在于隐蔽扫描主机、端口的行为。随机 IP 地址生成指的是对于要扫描的目标 IP 地址空间,生成一个完整随机不重复的序列,避免对目标进行顺序扫描,优点是增加了探测的隐蔽性,缺点是增加了计算随机 IP 地址序列的开销。多源 IP 扫描指的是使用多个源 IP 对目标进行扫描,优点是能够稀释单个 IP 探测被发现的风险,缺点是增加了探测成本。对比两种方法,从开销角度,随机 IP 地址生成增加了计算开销,多源 IP 增加了成本开销;从速度角度,随机 IP 地址生成增加了探测时间,多源 IP 扫描可以结合并行技术,速度更快。

机器学习技术主要通过使用相关算法对 TCP/IP 协议指纹数据进行训练以识别操作系统。其优点是带来了多种指纹特征识别的方法,准确率比传统指纹匹配高。其缺点是增加了运算开销,且不同研究选取的指纹不同,暂时还没有统一的标准。

4 未来研究方向

尽管网络资产探测技术已经有了众多科研成果,但要用于网络资产全生命管理还需要在以下三个方向进一步研究:

(1)探测结果的全面性

进行网络资产管理的前提就是要对被管理资产应知尽知,单个探测方法显然不足以达到全面探测的效果,一方面可以针对被忽略的网络资产研究新的探测方法,另一方面可以根据具体的场景结合多维探测技术以提高结果的全面性。

(2)探测结果的准确性

考虑到错误的探测结果可能不利于网络资产全生命周期管理的漏洞管理流程,需要知道正确的服务、操作系统版本才能根据该版本存在的漏洞进行预防、弥补工作。传统探测工具 Nmap、新型的机器学习方法在操作系统识别准确率方面依然存在不足,还需进一步研究。

(3)探测速度

对于提高探测速度的研究很多,但是在提高探测速度的同时,又会产生数据包的丢失问题,进而导致探测结果的全面性和准确性下降,如何在保证全面性和准确性的情况下尽可能提高探测速度也需要深入研究。

5 结论

网络资产探测在网络安全和网络资产管理方面是必不可少的部分,面对发展越来越快的网络技术和数量越来越多的网络资产,如何快速、全面、准确地进行网络资产探测也需要进一步研究。本文介绍了网络资产探测一些相关的技术,希望可以基于这些技术开发更优秀的网络探测工具,为学校、企业等组织的网络资产全生命周期管理提供便利。

参考文献

[1] 王宸东,郭渊博,甄帅辉,等.网络资产探测技术研究[J].计算机科学,2018,45(12):24-31.

[2] HAN X,DU X.A new method about operating system identification[C]//2010 2nd IEEE International Conference on Information and Financial Engineering,2010:882-885.

[3] JIANG W H,LI W H,JUN D.The application of ICMP protocol in network scanning[C]//Proceedings of the Fourth International Conference on Parallel and Distributed Computing,Applications and Technologies,2003:904-906.

[4] SONG J,KIM Y,WON Y.Operating system fingerprint recognition using ICMP[C]//Advances in Computer Science and Ubiquitous Computing,2018:285-290.

[5] LYON G F.Nmap network scanning:the official Nmap project guide to network discovery and security scanning[M].USA:Insecure,2008.

[6] DURUMERIC Z,WUSTROW E,HALDERMAN J A.ZMap:fast internet-wide scanning and its security applications[C]//22nd USENIX Security Symposium,2013:605-620.

[7] GRAHAM R.Masscan:the entire internet in 3 minutes[EB/OL].[2022-09-12].http://blog.erratasec.com/2013/09/masscan-entire-internet-in-3-minutes.html.

[8] GARCIA S.DNmap the distributed nmap[EB/OL].[2022-09-12].http://mateslab.weebly.com/dnmap-the-distributed-nmap.html.

[9] YUAN C,DU J,YUE M,et al.The design of large scale IP address and port scanning tool[J].Sensors,2020,20(16):4423.

[10] BEVERLY R.Yarrp′ing the Internet:randomized high-speed active topology discovery[C]//Proceedings of the 2016 Internet Measurement Conference, 2016:413-420.

[11] LI G, ZHANG M, GUO C, et al.Switches are scanners too! a fast and scalable in-network scanner with programmable switches[C]. Proceedings of the Twentieth ACM Workshop on Hot Topics in Networks, 2021: 77-83.

[12] LI G, ZHANG M, GUO C, et al.IMap: fast and scalable in-network scanning with programmable switches[C]// 19th USENIX Symposium on Networked Systems Design and Implementation, 2022: 667-681.

[13] ZHANG M, CHEN Y, CHEN H, et al. Design and implementation of a high performance network scanning system for VxWorks hosts[C]//2016 International Conference on Communications, Information Management and Network Security, 2016: 119-122.

[14] SIVANATHAN A, GHARAKHEILI H H, SIVARAMAN V. Can we classify an iot device using tcp port scan?[C]// 2018 IEEE International Conference on Information and Automation for Sustainability, 2018: 1-4.

[15] JUNG H C, JO H, LEE H. UDP-based active scan for IoT security(UAIS)[J].KSII Transactions on Internet and Information Systems(TIIS), 2021, 15(1): 20-34.

[16] KUMAR S, SUDARSAN S D. An innovative UDP port scanning technique[J]. International Journal of Future Computer and Communication, 2014, 3(6): 381.

[17] HANDOKO S, WARNARS H L H S. Network scanning topology based on inventory using query SNMP method[C]//Smart Data Intelligence, 2022: 295-306.

[18] SAHEB S I, RASOOL MD A. Auto-discovery and monitoring of network resources: SNMP-based network mapping and fault management[C]//Smart Computing Techniques and Applications, 2021: 643-653.

[19] ROQUERO P, ARACIL J. On performance and scalability of cost-effective SNMP managers for large-scale polling[J]. IEEE Access, 2021, 9: 7374-7383.

[20] 李憧, 刘鹏, 蔡国庆. 基于流量感知的动态网络资产监测研究[J]. 信息安全研究, 2020, 6(6): 523-529.

[21] ZALEWSKI M. p0f v3: passive fingerprinter[EB/OL]. [2022-09-12]. http://lcamtuf.coredump.cx/p0f3/README.

[22] ORNAGHI A, VALLERI M, ESCOBAR E, et al. The ettercap project[EB/OL]. [2022-09-12]. https://www.ettercap-project.org/.

[23] KOLLMANN E. Chatter on the wire: a look at DHCP traffic[EB/OL]. [2022-09-12]. http://myweb.cableone.net/xnih/download/chatter-dhcp.pdf.

[24] HJELMVIK. Networkminer homepage[EB/OL]. [2022-09-12]. https://sourceforge.net/projects/networkminer.

[25] BAI S, KIM H, REXFORD J. Passive OS fingerprinting on commodity switches[C]//2022 IEEE 8th International Conference on Network Softwarization(NetSoft), 2022: 264-268.

[26] LASTOVICKA M, JIRSIK T, CELEDA P, et al. Passive OS fingerprinting methods in the jungle of wireless networks[C]//NOMS 2018-2018 IEEE/IFIP Network Operations and Management Symposium, 2018: 1-9.

[27] ZOLOTYKH M. Research of behavior of the search engine 'Shodan. io'[C]//2020 Global Smart Industry Conference(GloSIC), 2020: 42-48.

[28] ZOLOTYKH M. Study of crawlers of search engine 'Shodan. io'[C]//2021 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology(USBEREIT), 2021: 0419-0422.

[29] 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.

[30] 胡栋梁, 秦晓军, 王晓锋. 基于消息中间件的分布式网络扫描[J]. 计算机工程, 2020, 46(12): 163-170.

[31] DERI L. Improving passive packet capture: beyond device polling[C]//Proceedings of SANE, 2004: 85-93.

[32] DERI L. PF_RING ZC(Zero Copy)[EB/OL]. [2022-09-12]. https://www.ntop.org/products/packet-capture/pf_ring/pf_ring-zc-zero-copy/.

[33] KIM J H, NA J C. A study on one-way communication using PF_RING ZC[C]//2017 19th International Conference on Advanced Communication Technology (ICACT), 2017: 301-304.

[34] YE J, LI J, JIANG A E P, et al. PF_RING-TA: a traffic-adaptive method of packet capturing by enhancing PF_Ring[C]//2021 IEEE 6th International Conference on Smart Cloud (SmartCloud), 2021: 51-56.

[35] LAŠTOVIČKA M, DUFKA A, KOMÁRKOVÁ J. Machine learning fingerprinting methods in cyber security domain: which one to use?[C]//2018 14th International Wire-

(下转第 35 页)

不可能一蹴而就，从企业评估自身能力开始，到组织机构变革创新、政策制定、流程重建等，都是较为详细的工作项目。从大数据分析的角度看，大数据治理缺少激动人心的业务创新，更多的是枯燥无味、苦练内功的持续投入。大数据治理工作的特点决定了企业大数据业务不可能迅速见效，领导层的决心和企业上下的协调一致是实现数据真正治理以及挖掘大数据价值的不二法门。

参考文献

[1] 大数据标准化白皮书[Z].全国信息技术标准化技术委员会大数据标准工作组，2020.
[2] 邹丹，马小宁，王喆.铁路大数据平台架构研究[J].铁路计算机应用，2019，28(8)：1-4.
[3] 郑大庆，范颖捷，潘蓉，等.大数据治理的概念与要素探析[J].科技管理研究，2017，37(15)：200-205.
[4] 甘似禹，车品觉，杨天顺，等.大数据治理体系[J].计算机应用与软件，2018，35(6)：1-8，69.
[5] 代红，张群，尹卓.大数据治理标准体系研究[J].大数据，2019，5(3)：47-54.
[6] 印鉴，朱怀杰，余建兴，等.大数据治理的全景式框架[J].大数据，2020，6(2)：19-26.
[7] 王俊，王修来，庞威，等.面向科技前瞻预测的大数据治理研究[J].计算机科学，2021，48(9)：36-42.
[8] 廖振民.大数据治理：传统政府治理的变革之道[J].桂海论丛，2018，34(2)：114-119.
[9] 赵豫生，林少敏，郑少翀.大数据治理机构职能及其评价指标体系构建研究[J].中国行政管理，2020(7)：70-77.
[10] 李冬，万磊，费建章.大数据治理中的安全问题研究[J].信息与电脑，2017(6)：192-193.
[11] 杨隆志，李洁，诺伊莉莎，等.网络安全中的大数据治理[J].信息安全与通信保密，2021(7)：56-66.

(收稿日期：2022-10-01)

作者简介：

王喆(1981-)，男，博士，副研究员，主要研究方向为云原生技术、大数据等。

(上接第 9 页)

less Communications & Mobile Computing Conference (IWCMC)，2018：542-547.
[36] FAN X，GOU G，KANG C，et al.Identify OS from encrypted traffic with TCP/IP stack fingerprinting[C]//2019 IEEE 38th International Performance Computing and Communications Conference(IPCCC)，2019：1-7.
[37] SONG J，CHO C H，WON Y.Analysis of operating system identification via fingerprinting and machine learning[J].Computers & Electrical Engineering，2019，78：1-10.
[38] KUMAR A，SONI I，ANAND KUMAR M.Operating system fingerprinting using machine learning[C]//Proceedings of International Conference on Intelligent Cyber-Physical Systems，2022：157-167.

(收稿日期：2022-09-12)

作者简介：

邵磊(1994-)，男，硕士，主要研究方向：网络安全。

余晓(1973-)，通信作者，女，硕士，讲师，主要研究方向：网络管理、云计算、网络安全。E-mail：pp_xyu@seu.edu.cn。

吴剑章(1972-)，男，硕士，讲师，主要研究方向：网络管理、物联网、云计算、虚拟化技术、机器学习，网络安全。