

防护性DNS服务关键技术研究综述

马永征, 李洪涛, 马中胜, 胡卫宏, 张中献

(中国互联网络信息中心, 北京 100190)

摘要: 域名系统 (domain name system, DNS) 作为互联网资源的名字标识服务, 提供了从域名到IP地址的查询转换功能, 是用户访问互联网应用的入口, 也是用户侧安全威胁感知与防御的关键点。用户侧常见的DNS滥用及相关安全威胁有: 网络钓鱼、域名不良应用、恶意软件以及利用DNS进行攻击等。防护性DNS (protective DNS, PDNS) 服务是一种利用DNS协议和架构的网络安全防护技术, 通过对DNS查询进行威胁检测与处置, 能够从源头上阻止用户对网络钓鱼网站、不良网站、恶意软件的访问。目前学界对PDNS服务还缺少系统的介绍与研究。对PDNS已有研究工作、应用现状及架构与功能进行系统梳理, 对PDNS所涉及的关键技术进行系统综述, 主要包括域名威胁处置技术、DNS异常检测技术、威胁情报管理技术和数据存储管理技术, 分析PDNS目前面临的问题与挑战, 并对PDNS未来发展趋势与研究方向提出了展望。

关键词: 域名系统; 防护性DNS服务; 域名威胁处置; DNS异常检测; 威胁情报管理; 数据存储管理

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2024065

Survey on key technologies for protective DNS service

MA Yongzheng, LI Hongtao, MA Zhongsheng, HU Weihong, ZHANG Zhongxian

China Internet Network Information Center, Beijing 100190, China

Abstract: The domain name system (DNS) is a naming service for Internet resources that provides the function of converting from domain names to IP addresses. It serves as the entry point for users to access Internet applications and is also the key point for client-side security threat perception and mitigation. Common DNS abuses and related security threats on the client side include phishing, malware, and other attacks based on DNS. The protective DNS (PDNS) service leverages the existing DNS protocol and architecture to detect and mitigate threats by analyzing DNS queries, thereby protecting users against potential security threats at the source. Currently, a systematic introduction and research on PDNS service in academia and industry is lacking. The existing work, application status, and architecture and functions of PDNS were systematically sorted out. The key technologies associated with PDNS, including DNS-related threat mitigation, DNS anomaly detection, threat intelligence management, and data storage and management technology, were reviewed. The issues and challenges faced by PDNS were analyzed, and

收稿日期: 2024-04-29; 修回日期: 2024-09-18

通信作者: 马中胜, mazhongsheng@cnnic.cn

基金项目: 国家重点研发计划 (2022YFB3105000)

Foundation Item: The National Key R&D Program of China (2022YFB3105000)

引用格式: 马永征, 李洪涛, 马中胜, 等. 防护性DNS服务关键技术研究综述[J]. 网络与信息安全学报, 2024, 10(5): 39-55.

Citation Format: MA Y Z, LI H T, MA Z S, et al. Survey on key technologies for protective DNS service[J]. Chinese Journal of Network and Information Security, 2024, 10(5): 39-55.

future development trends and research directions for PDNS were proposed.

Keywords: domain name system (DNS), protective DNS (PDNS) service, DNS threat mitigation, DNS anomaly detection, threat intelligence management, data storage management

0 引言

域名是一串字符标识,用于识别和定位互联网上的主机,与主机的IP地址相对应。域名系统(domain name system, DNS)作为互联网资源的名字标识服务,提供了从互联网域名到IP地址的查询转换服务,是用户访问互联网应用(如网页浏览、电子邮件和聊天等)所需要的一种基础服务,被视为互联网的入口。由于DNS在设计之初并未将安全作为首要考虑,因此多年来DNS一直面临着安全方面的威胁,也是各种网络攻击行为的重要针对目标。根据文献[1],2022年全球约有90%的组织受到DNS攻击,主要的攻击方式有网络钓鱼(54%)、基于DNS的恶意软件(49%)、分布式拒绝服务(distributed denial of service, DDoS)与放大攻击(35%)、DNS劫持、DNS隧道、缓存投毒等。其中DDoS、缓存投毒是针对DNS基础设施的攻击,其他则是由于DNS滥用对个人用户和企业公司等产生的安全威胁,后者可能对用户及企业造成不可估量的损失,近年来越来越受到重视和关注。

为了减轻用户侧DNS安全威胁的危害,业界采取了多种措施。部署方面,在网络关键节点部署DNS防火墙,主动检测恶意的域名解析请求,及时阻断用户对恶意网络应用的访问。协议方面,近年来陆续推出DoH(DNS over HTTPS)、DoT(DNS over TLS)、DoQ(DNS over QUIC)等安全传输协议,保障用户与DNS服务之间的数据传输安全,保护用户数据隐私。治理方面,对涉及违法及不良应用的域名进行清理。此外,ICANN等社群与组织也在域名滥用治理方面做出积极努力,其发起的DNS安全威胁缓解计划^[2](DNS security threat mitigation program)为社区组织提供了一个协作平台,基于各种与DNS安全威胁相关的倡议和项目形成集中战略以减轻DNS安全威胁。

随着DoH等安全加密传输协议的推广、应

用,传统DNS防火墙等用户侧安全防护措施可能面临失效风险,部署在递归解析服务的防护性DNS(protective DNS, PDNS)服务应运而生。作为新型用户侧DNS安全防护解决方案,近年来PDNS已被多个大型公共DNS递归解析服务支持,如Cloudflare、Quad9^[3]等,用于防止DNS滥用或攻击、过滤不良内容。同时,PDNS也在一些国家级DNS递归解析服务中得到重视与应用。英国国家网络安全中心(National Cyber Security Centre, NCSC)关于PDNS部署发布了官方建议;美国网络安全和基础设施安全局(Cybersecurity and Infrastructure Security Agency, CISA)向联邦民事行政部门(FCEB)机构提供了PDNS服务;欧盟委员会发起DNS4EU倡议^[4],作为欧盟新网络安全战略的一部分。防护性DNS服务在全球呈现出良好的发展态势。

近年,学术界和业界也对PDNS开展了研究。文献[5]介绍了英国NCSC的PDNS服务,展示了PDNS如何对DNS功能进行补充,以提高应用机构的安全性和隐私性。其案例研究表明,PDNS服务可作为深度防御策略的一部分,能够在域名解析时阻止恶意内容访问。文献[6]针对大众用户、互联网服务提供商(Internet service provider, ISP)用户、企业专业人员、DNS技术专家4类人群,进行了关于PDNS的研究,研究表明,用户使用PDNS的原因主要有:增强安全、获取全球威胁情报、遵守政府法规。文献[7]提出PDNS服务探测识别方法,对常用的PDNS服务提供商进行了分析,总结了DNS重写策略的主要形式和PDNS的部署情况。PDNS是涉及域名系统、网络安全、大数据、网络管理等综合性服务系统。这些文献仅研究了PDNS的功能、部署等某些方面,对其关键技术还缺少综合性和系统化的研究。

本文详细阐述了PDNS的概念与功能、体系结构以及发展现状等,并对PDNS所涉及的关键技术进行梳理研究,最后总结了PDNS的未来发展前景和面临的挑战。

1 背景

1.1 域名系统

域名系统是一种层级的分布式系统，它主要用于将人类易读的名字（域名）映射到网络地址（即IP地址），映射信息以资源记录的形式存储在分布式数据库中。RFC 1034^[8]中描述了DNS的标准查询方式。如果要从DNS数据库中查询信息，客户端通过解析器（递归解析服务器）进行查询，在请求中指定域名以及要获取的资源记录（resource record, RR）类型。解析器向权威解析服务器进行迭代查询，并将DNS响应返回给客户端。一旦解析器获得了给定域名的IP地址，将对响应信息进行缓存。缓存信息在解析器收到客户端的查询请求时，用于检查是否存在请求所对应的应答。

DNS的解析主要由客户端（用户侧）与服务端（DNS基础设施）构成，如图1所示。相应地，DNS安全问题可分为用户侧DNS安全与DNS基础设施安全。PDNS主要用于用户侧DNS安全防护，因此本文的重点研究范围为用户侧DNS安全威胁与问题。

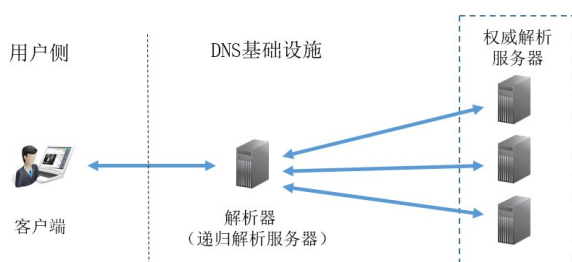


图1 DNS解析主要构成示意

Figure 1 Overview of components in DNS resolution

1.2 用户侧DNS安全威胁与问题

用户侧DNS安全一般是指用于保护用户在DNS解析过程中免受恶意行为影响的技术或策略。用户侧DNS解析的安全威胁和问题主要包括：网络钓鱼（phishing）、域名不良应用、恶意软件攻击、速变恶意域名、字符注入攻击、DNS重绑定攻击等。

1.2.1 网络钓鱼

网络钓鱼这一术语是由钓鱼（fishing）一词演变而来，是利用社会工程学和技术手段窃取消费者的个人身份数据和财务账户凭证的网络攻击

方式。攻击者往往是向用户发送貌似来自合法企业或机构的欺骗性电子邮件、手机短信等，引诱用户回复个人敏感信息或点击里面的链接，访问伪造的网站^[9]，进而泄露凭证信息（如用户名、密码）或下载恶意软件。还有一些网络钓鱼攻击，攻击者直接在计算机上移植恶意软件（如浏览器中间者攻击），采用某些技术手段直接窃取凭证信息，如使用系统拦截用户的用户名和密码、误导用户访问伪造的网站等。

1.2.2 域名不良应用

域名被用于色情、网络赌博等违法、违规内容的网站，属于域名不良应用。根据中共中央网络安全和信息化委员会办公室（中央网信办）违法和不良信息举报中心统计^[10]，2022年全国受理网络违法和不良信息举报1.72亿件，主要包括色情、赌博、侵权、谣言等违法和不良信息。

1.2.3 恶意软件攻击

恶意软件（malware）^[11]是对各类干扰计算机正常运作、收集敏感信息、入侵计算机系统的程序的统称，包括勒索软件、蠕虫、特洛伊木马、僵尸网络或其他有恶意目的的程序。恶意软件在网络通信时，往往利用DNS来进行机密信息窃取、组建僵尸网络。僵尸网络^[12]是攻击者基于命令和控制（command and control, C&C）服务器组建起来的网络，常被用于发起DDoS攻击或者发送垃圾邮件。

1.2.4 速变恶意域名

在恶意软件中，常会使用速变（flux）技术进行IP地址或域名的快速切换，来连接控制服务器。速变技术主要分为：Fast-Flux（也被称为IP-Flux）和Domain-Flux。Fast-Flux技术是指不断改变域名同IP地址映射关系的一种技术，为了避免僵尸网络中的C&C服务器被检测到，攻击者使用Fast-Flux将僵尸网络控制的主机组建成为具有负载均衡和高可用性的代理网络，提高僵尸网络的健壮性。Domain-Flux技术，如域名生成算法（domain generation algorithm, DGA）可以生成大量的随机域名，使C&C服务器的地址难以被检测或阻断。

1.2.5 字符串注入攻击

在字符串注入攻击^[13]中，攻击者将恶意载

荷信息在权威解析服务器的 DNS 记录中进行编码, 通过 DNS 隧道向 DNS 递归解析服务器返回带有恶意代码的 DNS 记录, 进而将恶意代码注入用户终端。字符串注入攻击能够导致系统崩溃、数据损坏和泄露、远程代码执行等风险。一些 DNS 隧道工具 (如 Iodine) 的开源, 增加了攻击者使用工具进行字符串注入攻击的威胁。

1.2.6 DNS 重绑定攻击

DNS 重绑定^[14]是一种绕过浏览器同源策略的限制, 并将浏览器作为代理的攻击方式。DNS 重绑定倾向于控制内部网络中的物联网 (Internet of things, IoT) 设备或者窃取敏感文件^[15]。攻击者通过欺骗用户将其访问的域名映射到恶意 IP 地址上, 以达到绕过浏览器同源策略渗透进内网, 从而控制设备及窃取敏感信息的目的。攻击者创建一个恶意网站, 并将其托管在正常的域名下。用户在访问这个域名时, 首次解析的请求会返回合法的 IP 地址。一旦用户与该恶意网站建立了连接, 在网站的响应中会有恶意代码, 使后续的解析切换到恶意的 IP 地址上。

1.3 用户侧 DNS 安全协议扩展

为保障 DNS 数据在传输过程中不被更改, IETF 提出了 DNS 安全扩展 (domain name system security extensions, DNSSEC) 技术体系。DNSSEC 技术体系利用公钥基础设施, 在 DNS 原有的体系结构上添加数字签名, 为 DNS 消息提供了来源认证和信息完整性验证, 能够保障 DNS 数据来源的准确性, 确保数据在传输过程中不被篡改, 同时也能够为否定应答报文提供验证。

随着网络规模爆发式增长和网络环境的不断变化, 域名劫持、记录篡改、大规模监控及用户隐私泄露等问题频发, 推动技术社群越来越多地采用隐私保护和加密认证技术增强 DNS 的安全性。为了增强传输过程中的隐私保护, IETF 先后制定了 DoH、DoT 和 DoQ 等 DNS 信道传输加密协议, 旨在从隐私保护、身份认证、数据完整性等方面增强 DNS 数据的安全性并最大限度地减少网络运营商等对 DNS 流量的监控分析。DoH 使用 HTTPS 进行 DNS 消息数据的传输。DoT 通过 TLS (transport layer security) 连接发送

DNS 数据, 利用 TLS 实现 DNS 消息的加密和认证。DoT、DoH 技术已经得到越来越多大型互联网企业的部署和应用, 但也存在一些问题, 包括建立连接的握手延迟、队头阻塞等。DoQ 协议在上述背景下被提出, 将 DNS 请求和响应消息在 QUIC (quick UDP Internet connections) 报文中进行封装, 利用 QUIC 的安全特性为 DNS 提供隐私保护等功能。

DNSSEC 技术体系与 DNS 安全传输技术对 DNS 协议和架构进行了扩展和增强。然而, DNSSEC 技术体系依赖于递归解析服务与权威解析服务的同步部署。DoH 等加密传输技术在实现隐私保护的同时, 也产生了利用隐蔽通信和绕过安全策略的风险。例如, 攻击者可以利用 DoH 绕过网络防火墙, 访问恶意网站或执行钓鱼攻击。此外, DoH 也可能被用作恶意软件的通信通道, 由于 DNS 查询和响应内容的不可见性 (即无法知道 DNS 查询或响应中特定字段的内容), 大多数现有的基于报文特征检测 DNS 隐蔽通道的方法将失效。

综上, 随着 DNS 安全传输技术逐渐被广泛应用, 在递归解析服务中部署 PDNS, 可以在一定程度有效减小用户侧 DNS 的安全威胁, 而且不依赖于 DNS 协议扩展相关的部署, 是网络安全领域具有潜力的一个发展方向。

2 PDNS

2.1 PDNS 介绍

PDNS^[3]是一种利用现有 DNS 协议和架构, 通过分析 DNS 查询并采取措施来减小用户侧安全威胁的服务。图 2 给出了 PDNS 示意。PDNS 可以记录和保存可疑的 DNS 查询, 并通过修改应答内容, 实现延迟或防止恶意行为的效果, 同时相关机构或组织也能够对这些可疑的 DNS 查询进行进一步调查。PDNS 不需要更改 DNS 协议, 可以从源头上阻止用户侧具有安全威胁的域名访问, 从而增强网络安全。

2.2 PDNS 现状

目前全球已有很多 PDNS 服务部署, 根据服务提供方以及服务目标对象的不同, PDNS 服务可分为两种类型: 国家或区域级 PDNS 服务与域名服务商的 PDNS 服务。

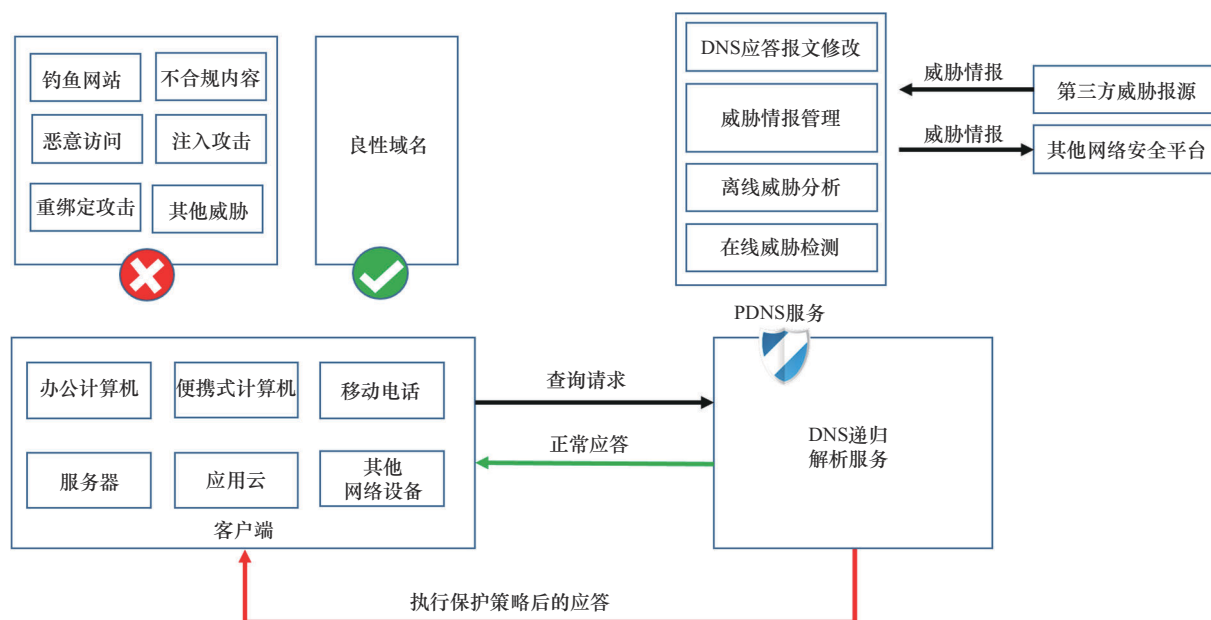


图2 PDNS示意

Figure 2 Diagram of PDNS

2.2.1 国家或区域级PDNS服务

比较典型的国家或区域级PDNS服务有美国的网络分析和数据系统 (cyber analytic and data systems, CADS)、英国NCSC的PDNS服务、欧盟的DNS4EU计划、加拿大互联网注册管理局 (Canadian Internet Registration Authority, CIRA) 的“加拿大盾”服务。

美国CISA的CADS计划继承了“爱因斯坦计划”中分析、信息共享和核心基础设施的能力，能够快速地分析、关联并采取行动应对网络安全威胁^[16]。其PDNS服务分布在不同的地点，获取来自商业技术提供商、政府机构和其他渠道的威胁情报信息，能够对DNS流量进行分析检测，实行减小流量中网络威胁的措施。

英国NCSC创建并管理主动网络防御 (active cyber defence, ACD) 解决方案，PDNS服务是ACD解决方案的能力之一。英国NCSC的PDNS服务能够阻止对已识别的恶意网络链接进行DNS解析。该服务被提供给英国的公共机构，以保护整个英国的中央政府部门和其他公共部门组织的安全。

欧盟委员会于2020年年底宣布了一项新的网络安全战略，包括一系列新的网络安全举措。新战略的要素之一是DNS4EU提案^[17]。DNS4EU旨在建立欧洲公共DNS递归解析服务，防止该

领域被其他外部的公共DNS解析服务垄断。2022年，欧盟委员会启动了DNS4EU计划^[6]，建设欧洲PDNS服务，以满足欧盟对DNS的弹性、安全和隐私等方面的需求，并且针对恶意软件和其他网络威胁提出了明确的过滤政策。

加拿大CIRA推出的“加拿大盾”为普通网民提供支持PDNS的公共DNS递归解析服务。服务支持多种模式：无过滤的加密连接、阻止恶意软件和网络钓鱼、网络安全保护与色情内容阻止。“加拿大盾”的网络威胁情报主要来自开源社区、商业技术提供商，以及加拿大网络安全中心的威胁情报数据^[4]。

2.2.2 域名服务商的PDNS服务

荷兰域名“.NL”的注册管理机构 (Netherlands domain registration foundation, SIDN) 公司的ENTRADA (enhanced top-level domain resilience through advanced data analysis) 平台提供了PDNS相关的服务。ENTRADA平台^[18]能够构建安全的应用程序来检测恶意活动，如网络钓鱼攻击、检测僵尸网络域名，并将这些威胁情报提供给托管提供商和互联网服务提供商，以帮助对抗网络威胁。

递归服务方面，目前业界对PDNS概念及服务能力的界定尚无统一的标准，因此，判断DNS服务是否支持PDNS功能还有一定难度。文献[7]

根据常用 DNS 服务提供商的官方描述, 粗略统计得出约 67% 公共递归服务提供 (部分) PDNS 功能, 国内的有 OneDNS、阿里云 DNS、360 DNS、114 DNS、DNSPod 等, 国外的有 Quad9、Cloudflare、Google 等。该文献还对 DNS 解析服务的 PDNS 能力进行探测, 结果显示仅有少数 DNS 服务提供了 PDNS 类的功能, 占被探测服务器的 9.1%。

2.3 PDNS 功能

PDNS 服务利用现有 DNS 协议和架构来分析 DNS 查询, 并基于响应策略采取措施以应对恶意或可疑的 DNS 查询。PDNS 服务能够从源头上阻止对恶意软件、勒索软件、网络钓鱼攻击、病毒、恶意站点和间谍软件的访问, 从而使用户免受其危害, 增强了网络安全性。例如, 针对一些欺诈和不良网络应用, PDNS 可以过滤掉域名解析请求, 从而防止用户访问不良内容网址。PDNS 服务使用的已知恶意域名列表信息通常来自开源、商业的或政府信息源等。此外, 在一些应用中, PDNS 可以基于异常检测技术来检测新的恶意域名。PDNS 通常能够识别处理的域名威胁类型如下: 钓鱼网站、恶意软件分发、C&C 行为、DGA、不良内容等。典型的 PDNS 服务的核心功能主要有: 域名威胁情报管理、域名威胁处置、域名威胁感知以及数据存储管理。

(1) 域名威胁情报管理

威胁情报一般包括漏洞库、IP 信誉库、工具库、DNS 库、统一资源定位符 (uniform resource location, URL) 库等, 并基于外界的情报源进行动态、实时的更新, 确保情报的有效性。PDNS 可整合多个来源的威胁情报数据, 包括来自外部和内部的威胁信息。外部的威胁信息可来自政府或行业合作伙伴的非机密商业威胁情报数据。内部的威胁信息是在系统内通过主动识别检测生成的, 并且可以共享给外部相关安全机构, 用于网络安全防护策略制定和实施。

(2) 域名威胁处置

PDNS 支持针对不同威胁设置不同的处置策略。例如, 针对钓鱼网站, 防止用户意外连接到有潜在危害的恶意网址; 针对恶意软件, 阻止已

知的恶意连接尝试; 针对 DGA, 分析和标记与已知的 DGA 属性相关的域名; 针对不良应用域名, 对不良内容进行过滤, 给出警告或阻止访问。由于 PDNS 系统可以面向涉及 DNS 阻断、内容过滤、攻击拦截等不同防护需求的用户, 用户的 DNS 防护策略存在差异。因此, PDNS 需要能够支持异构的防护策略, 以满足不同的 DNS 防护需求。

(3) 域名威胁感知

域名威胁感知是利用黑白名单、智能分析等检测识别技术, 提取域名解析流量特征和指标, 发现域名安全威胁。威胁检测方式分为离线分析和在线检测。威胁检测针对不同威胁类型有钓鱼网站检测、恶意软件检测、DGA 检测、不良域名应用检测等。为了避免误识别, 对已识别的 DNS 安全威胁, 还需要进行进一步的威胁鉴定。

(4) 数据存储管理

PDNS 需要具备海量数据存储能力, 存储 DNS 解析日志、DNS 查询记录、PDNS 处置记录等数据。相关的数据可供后续审计或分析。数据存储前, 涉及数据采集、数据处理等环节。数据采集关注采集的数据类型以及采集方式。数据处理主要关注如何处理采集到的数据, 以及将采集到的数据进行有效融合。

3 PDNS 关键技术

根据 PDNS 的功能分类, PDNS 关键技术可分为 4 类: 域名威胁处置技术、DNS 异常检测技术、域名威胁情报管理技术、数据存储管理技术。PDNS 功能与关键技术的对应关系如图 3 所示。域名威胁处置技术包括响应策略区 (response policy zones, RPZ) 技术和 DNS Sinkhole 技术; DNS 异常检测技术包括网络钓鱼检测技术、域名不良应用检测技术、恶意软件攻击检测技术、速变域名检测技术、DNS 隧道攻击检测技术、DNS 重绑定攻击检测技术、恶意域名检测与分类技术; 域名威胁情报管理技术包括黑名单与网络威胁情报 (cyber threat intelligence, CTI) 技术; 数据存储管理技术包括数据仓库、数据湖、湖仓一体技术。

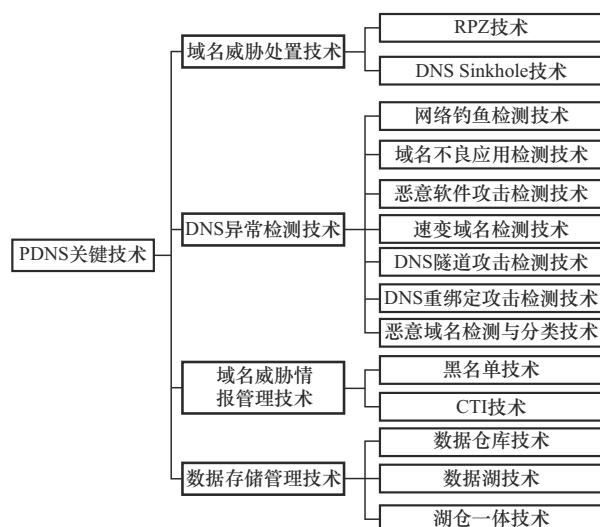


图3 PDNS功能与关键技术的对应关系

Figure 3 Correspondence of PDNS functions and key technologies

3.1 域名威胁处置技术

3.1.1 RPZ技术

RPZ在文献[19]中被提出,它可以用于预防DNS相关威胁的发生。RPZ通过预先制定的策略,阻止DNS查询客户端访问恶意域名或者IP,进而实现阻止客户端访问受感染或者恶意站点。在策略制定时,可以为特定的DNS查询(或响应)创建对应的策略。策略内容包括触发的条件(“触发器”)和需要采取的行动(“动作”)。

RPZ的“触发器”有5种,包括客户端IP地址(Client IP Address)、查询的域名(QNAME)、响应IP地址(Response IP Address)、名称服务器域名(NSDNAME)、名称服务器IP地址(NSIP)。RPZ的“动作”有6种,包括回复域名不存在(NXDOMAIN)、回复记录类型不存在(NODATA)、透传(PASSTHRU)、丢弃查询请求(DROP)、TCP再次发起域名查询(TCP-ACTION)、回复本地预先配置的解析结果(LOCAL DATA)。主流DNS软件如BIND、Unbound和PowerDNS Recursor等,均已支持基本的RPZ功能。

文献[20]介绍了美国关于RPZ的应用实践,对启用RPZ功能的公共DNS服务的阻止行为与其他DNS服务的阻止行为进行了比较,结果显示,启用RPZ对恶意域名有显著的阻止效果。文献[21]针对僵尸网络中受控计算机违反DNS协议

流程的安全问题,提出了一个基于策略的检测和阻止系统,使用RPZ来阻止异常DNS查询。

3.1.2 DNS Sinkhole技术

DNS Sinkhole通过拦截和重定向DNS请求来工作,旨在对抗和中和恶意的在线活动。针对僵尸网络,DNS Sinkhole通过阻止C&C服务器与僵尸网络中被控制计算机之间的通信,来减小被控制计算机的攻击活动。被控制计算机在向DNS服务器发送DNS查询以获取C&C服务器IP地址时,针对来自C&C服务器域名黑名单的请求,DNS服务器以DNS Sinkhole服务器的IP地址进行响应,使被控制计算机与DNS Sinkhole服务器进行通信,进而切断被控制计算机与C&C服务器之间的通信,避免受C&C服务器控制。

DNS Sinkhole还可用于阻止基于垃圾邮件的攻击。文献[22]将没有有效发件人和收件人地址的退回电子邮件作为垃圾邮件,从中提取恶意URL并将其应用到DNS Sinkhole系统中。

3.2 DNS异常检测技术

PDNS相关的DNS安全检测归属于网络异常检测领域,识别的异常类型主要有网络应用异常、网络行为异常、网络流量异常。主要针对网络钓鱼、域名不良应用、恶意软件攻击、恶意域名、DNS重绑定攻击、DNS隧道攻击等场景。检测对象主要包括网站入口、网站内容、网络数据流。针对DNS对应网站的滥用(网络钓鱼、域名不良应用),检测对象主要是网站入口与网站内容;针对恶意软件攻击、速变恶意域名、DNS重绑定攻击、DNS隧道攻击,检测对象主要是网络数据流。此外还有一些DNS异常检测的检测对象为解析日志。

DNS异常检测的方法主要有两类:基于规则匹配和基于模型识别。前者如基于黑白名单的检测,触发预设规则后发生匹配;后者包括基于统计的检测、基于机器学习的检测、基于深度学习的检测等。表1列出了不同DNS异常检测类型的检测对象与检测方法。

基于黑白名单的检测,需要维护一个已知的异常域名、IP或网站信息列表,以便根据列表对当前访问的域名、IP或网站进行检查。

基于模型识别的异常检测方法中,统计的方

表 1 不同 DNS 异常检测类型的检测对象与检测方法

Table 1 Objects and methods of different DNS anomaly detection

检测类型	检测对象	检测方法
网络钓鱼检测	网站入口、网站内容	黑白名单、统计、机器学习、深度学习
域名不良应用检测	网站入口、网站内容	黑白名单、统计、机器学习、深度学习
恶意软件攻击检测	网络数据流	黑白名单、统计、机器学习、深度学习
速变域名检测	网络数据流	统计、机器学习、深度学习
DNS 隧道攻击检测	网络数据流	规则匹配、机器学习、深度学习
DNS 重绑定攻击检测	网络数据流	统计、机器学习、深度学习
恶意域名检测与分类	网络数据流	黑白名单、机器学习、深度学习

法主要有参数方法和非参数方法。根据学习形式，机器学习主要分为有监督学习、无监督学习和半监督学习。有监督学习依赖于数据中的标签，通常需要大量数据进行模型的训练，方法有：支持向量机（support vector machine, SVM）、随机森林、J48 决策树、朴素贝叶斯。无监督学习事先不需要训练样本，检测性能通常低于有监督学习，方法有：K 均值、层次聚类等。半监督学习需要训练少量样本，利用大量未标记的样本实现异常的检测与识别。基于深度学习的方法主要有：循环神经网络（recurrent neural network, RNN）、长短期记忆（long short-term memory, LSTM）、卷积神经网络（convolutional neural network, CNN）等。这类方法的优点在于它可以利用序列和结构信息自动提取特征，这也意味着影响深度学习模型性能的具体原因可能未知，并且对硬件的计算性能要求很高。

3.2.1 网络钓鱼检测技术

绝大多数的网络钓鱼引诱用户访问其事先搭建好的仿冒网站。为了更好地获得用户的信任，钓鱼攻击者构建的钓鱼网页往往与真实网页非常相似^[14]，这种相似性包括 LOGO（机构标志图像）、网页图标（favicon）、层叠样式表（cascading style sheet, CSS）架构、布局及网页整体视觉的相似性。网络钓鱼检测的对象主要有：网站入口和网站内容。网站内容主要包括：文本内容和图片内容。一些研究也在关注多模态特征融合算法，如域名的语义特征、网页的文本特征^[23]、LOGO^[24]等。根据检测手段，网络钓鱼检测技术

可以分为：基于黑白名单的方法、基于统计的方法、基于机器学习的方法、基于深度学习的方法。

（1）基于黑白名单的方法

在基于黑白名单的检测方法中，需要维护一个已知的钓鱼网站信息列表，以便根据列表对当前访问的网站进行检查，将可疑 URL 与黑名单中的网络钓鱼 URL 进行匹配。这份列表需要不断更新，包括已知网络钓鱼的 URL、IP 地址、域名、证书或者关键词信息。

（2）基于统计的方法

基于统计的方法有启发式钓鱼检测和基于视觉相似性的钓鱼检测。启发式钓鱼检测根据网络钓鱼之间的相似性，从已检测到的网络钓鱼攻击中提取出特征，基于识别的特征进行网络钓鱼判断。由于启发式的规则特征主要来自网络钓鱼的统计特征或人工总结，该类方法的规则更新困难，存在误检率高的问题。基于视觉相似性的钓鱼检测针对网页内容进行判断，通过比较页面之间的视觉特征来实现网络钓鱼检测。文献[25]提出了基于 HTML 文本匹配的相似性度量检测方法，通过基准网页和潜在钓鱼网页的标签序列树对阈值来检测网络钓鱼。文献[26]对搜索结果中可疑的网站和匹配域名的网页内容进行比较，计算它们之间的相似度分数，作为钓鱼网站的检测参数。

（3）基于机器学习的方法

如果检测对象为文本，网络钓鱼检测问题可以视为一个文本分类或聚类的问题，可以运用机

器学习中分类算法、聚类算法进行检测。文献[27]针对URL提取了14个特征,使用SVM和朴素贝叶斯分类器进行网络钓鱼检测。文献[28]基于词典、WHOIS信息、网页排名、流量排名和页面重要性等特征属性进行网络钓鱼检测,使用SVM、KNN、决策树、随机森林等机器学习分类技术。

(4) 基于深度学习的方法

深度神经网络常被用于学习数据中的深层隐藏信息。文献[29]针对网络钓鱼设计了两类特征(原始特征和交互特征),提出了基于深度信念网络(deep belief network, DBN)的检测模型,用来进行网络钓鱼检测。

3.2.2 域名不良应用检测技术

网络赌博、淫秽色情等不良域名网页内容存在高度的相似性。目前针对不良网站的检测方法主要包括:基于黑白名单、基于统计、基于机器学习、基于深度学习的方法。

(1) 基于黑白名单的方法

首先建立黑白名单,通过网站的IP地址、URL、代理商等信息进行不良网站检测。这种检测方式准确率高,但存在滞后性,对于伪造IP地址、更换URL等手段不能很好地处理。

(2) 基于统计的方法

这类方法主要基于文本相似性与图片相似性进行网站内容检测^[30]。文本检测方面,随着网络信息指数级增长,敏感词越发多变,现有的文本过滤筛查机制准确率也相应下降。图片检测方面,随着图片样本空间的增大,不良图片隐蔽性越来越强,现有的图片相似性算法容易出现误判、错判、无法判定等现象。文献[30]根据大量不良域名内容高度相似性特点,提出了快速发现和处置“涉黄”“涉赌”两类网站的系统方案,在不良文本识别环节,通过建立文档主题的概率分布模型来提取关键词。

(3) 基于机器学习的方法

文献[31]从URL中提取文本特征,分别使用SVM、随机森林、对率回归、朴素贝叶斯、多层感知器(multilayer perceptron, MLP)等多种机器学习算法来处理和分析数据,识别和分类与非法在线赌博相关的URL。文献[32]提出

了一个基于视觉和文本内容的自动化检测系统进行色情和赌博网站检测,系统从网页中提取文本特征与图片特征,对文本特征使用SVM算法进行分类,对图片特征使用随机森林算法进行分类,再通过对率回归算法进行数据融合与决策。

(4) 基于深度学习的方法

文献[33]使用预训练的Word2Vec词向量作为初始嵌入层,使用双向长短期记忆(Bi-LSTM)算法训练文本分类器,提取屏幕截图中文本内容的语义特征,对赌博网页进行检测。

3.2.3 恶意软件攻击检测技术

恶意软件利用域名发起的攻击有:蠕虫攻击、垃圾邮件攻击、僵尸网络攻击等。针对恶意软件攻击进行检测的数据源通常为网络流量,主要分为:基于全报文网络流量的检测和基于DNS流量的检测。

基于全报文网络流量的恶意软件检测主要基于网络数据流工具,如NetFlow、Snort等。文献[34]利用NetFlow、sFlow等流捕获技术检测僵尸网络活动。文献[35]基于网络入侵检测系统Snort,检测恶意软件感染过程的特定阶段,由入站扫描、漏洞利用、插件下载、控制通信、对外攻击等行为总结出恶意软件的活动轨迹。

与全报文网络流量相比,DNS流量的数据量相对较小,可以有效缓解性能压力。另外,由于域名解析是网络服务的入口,通过DNS数据可以在攻击活动开始之前完成检测。文献[36]提出了利用DNS数据进行僵尸网络恶意域名识别的系统Segugio,可部署到不同的ISP网络中,用于检测ISP网络被恶意软件感染的机器。

一些研究针对特定的恶意软件攻击进行检测。文献[37]研究了电子邮件蠕虫对DNS查询流量特征的影响,提出了基于无监督学习和时间序列分析邮件蠕虫检测方法。文献[38]基于垃圾邮件发送者和良性域名具有不同的发件人策略框架(sender policy framework, SPF)规则和不同的流量配置文件的特点,构建了一种新方法,使用从被动DNS流量收集的特征来检测垃圾邮件。文献[39]通过DNS流量来检测僵尸网络,查看DNS流量中的群组活动,实现对僵尸网络服务器

迁移的识别。文献[40]提出了一种快速僵尸网络检测算法,利用查询成功域名和查询失败域名之间的时间相关性和域名信息熵的相关性来提升检测速度。

3.2.4 速变域名检测技术

恶意软件大多会利用动态 DNS 解析技术实现受感染机与 C&C 服务器之间的通信,如 Fast-Flux 和 Domain-Flux。针对速变域名的检测技术可以分为:Fast-Flux 检测和 Domain-Flux 检测。

(1) Fast-Flux 检测

Fast-Flux 是一种 DNS 技术的应用,对 DNS 的 RR 设定非常短的生存时间 (time to live, TTL),并在一个 IP 集合中做频繁替换。这些 IP 表示受控主机集合,在 Fast-Flux 技术中作为“代理”角色。Fast-Flux 检测方式有主动 DNS 探测和被动 DNS 分析;检测的主要方法有机器学习方法和深度学习方法。

1) 机器学习方法。文献[41]选取域名属性、网络属性、IP 分布 3 个维度的 6 个特征项,基于加权 SVM 方法进行检测。文献[42]提出的 Flux-Buster 系统通过单链接层次聚类算法对域名进行聚类,生成相关的域名簇和 IP 地址簇,每个簇是潜在的 Fast-Flux 网络。文献[43]从 DNS 响应、IP 地址地理分布、联网设备搜索引擎相关这 3 个维度选取了 8 个主要特征,对 MLP、径向基函数 (radial basis function, RBF) 网络、SVM 这 3 种机器学习算法进行了比较, SVM 算法具有较高的检测精度。

2) 深度学习方法。文献[44]将域名特征分为空间特征、文本特征和侧信息特征 3 类,定义 DNS 响应包大小、响应延迟等信息为域名的侧信息,然后分别使用图卷积网络 (graph convolutional network, GCN)、Bi-LSTM、MLP 方法提取特征,并利用神经网络将几种特征进行融合,完成检测。

(2) Domain-Flux 检测

DGA (domain generation algorithm) 是 Domain-Flux 采用的主要技术,已被广泛应用于恶意软件、僵尸网络、勒索软件等。DGA 通过算法生成随机字符序列,并将这些字符序列用作

域名。由于这些域名看起来是随机生成的,因此不容易被传统的基于黑名单的防御系统识别。恶意软件或僵尸网络利用这些域名建立与 C&C 服务器的连接,从而接收指令或发送收集到的数据。基于 DGA 域名的特点对数据进行分析,是 DGA 域名检测的主要途径。Domain-Flux 检测方法主要有:统计的方法、机器学习方法、深度学习方法。

1) 统计的方法。文献[45]提出了使用字符分布频率分析和域名加权分数的方法来检测 DGA。文献[46]通过统计来自主机的 NXDOMAIN (域名不存在) 请求来进行 DGA 检测,这种方法不依赖于训练数据,也不受特定域名特征约束。

2) 机器学习方法。文献[47]使用 SVM 模型进行 DGA 检测,提取了 5 个属性:域名长度、域名熵、域名中元音比例、域名中连续辅音比例、域名中数字比例。文献[48]采用分类分析和聚类分析两阶段检测法实现了一种 DGA 域名检测系统,基于随机森林算法进行分类分析,基于 X 均值算法进行聚类分析。

3) 深度学习方法。文献[49]基于 LSTM 和 CNN 的模型,使用深度学习方法对 DGA 域名进行分类检测。文献[50]使用 LSTM 方法实现了 DGA 恶意域分类器,并使用对抗性机器学习来生成前所未见的恶意软件域名样本,用新样本扩充训练数据集,从而使机器学习模型的训练能更有效地检测到新的恶意域名变体。文献[51]结合了安全多方计算和差分隐私两种隐私增强技术,使用一维 CNN、LSTM 和 MLP 这 3 种模型进行 DGA 检测。

3.2.5 DNS 隧道攻击检测技术

隧道攻击检测主要有载荷和流量两个分析维度^[52-53]。载荷分析是针对一个或多个 DNS 数据包,其重点是分析载荷特征,如请求和响应的长度、主机名的熵、特定 DNS 隧道工具的签名、请求域名相关的功能等。流量分析侧重于 DNS 查询的整体流量,关注流量的异常变化,从更宏观的角度来分析 DNS 隧道的特点。流量特征有:每个 IP 地址的 DNS 流量、每个域名的 DNS 流量、每个域名的主机数、DNS 服务器的地理位

置、域名的历史信息、NXDOMAIN域名的响应数量等。DNS隧道攻击的检测方法主要有:基于规则、基于机器学习和基于深度学习的方法。

(1) 基于规则的方法

通过分析DNS隧道流量的特征表现,进而得出某些规则,被采集的DNS流量表现出的特征如果与这些规则相吻合,则会被判定为恶意隧道流量。基于规则进行隧道攻击检测的重点在于规则的设计,主要的设计方法有:基于签名的方法和基于阈值的方法。签名是指在数据包中发现的能够代表恶意隧道的特定字节。阈值是能够用于检测DNS隧道流量的特征阈值。DNS隧道攻击中,通常会尝试将尽可能多的数据放入请求和响应中,因此,DNS请求及响应的载荷大小是一个重要指标。文献[54]通过对DNS流量进行分析,得到用于Snort的规则,来检测DNS隧道。文献[55]提出了基于字符频率的域名指纹研究,分析DNS查询和响应中域的一元、二元和三元字符频率,根据字符频率的排序以及变化,来区分良性域名和隧道攻击域名。

(2) 基于机器学习的方法

文献[56]提取了用于区分合法查询与隐蔽通信的12个数据分组特征,使用J48决策树、对率回归和朴素贝叶斯3种算法模型,利用分类器进行隧道攻击判别。文献[57]选取了时间间隔、请求数据包大小、记录类型和子域熵4种特征,基于对率回归、SVM和决策树3种算法,构建检测模型。文献[58]提出了基于会话行为和随机森林算法的自动检测方法,选择9个特征来识别DNS隧道攻击行为。

(3) 基于深度学习的方法

CNN是常用的DNS隧道检测深度学习算法。文献[59]提出了一种基于CNN的检测方法,在模型训练之前,将每个DNS数据包转换为固定长度的字节表示向量,预处理后的数据被用于训练CNN模型。文献[60]结合深度学习和集成学习的思想,构建了一个由单模型检测决策器和多模型检测决策器组成的系统,使用的深度学习算法包括:密集神经网络(dense neural network, DNN)、一维CNN、RNN。

3.2.6 DNS重绑定攻击检测技术

DNS重绑定攻击在域名名称、时间、异常通信及恶意行为等方面存在特征^[61]。通常,DNS重绑定使用的域名是数字、字母和特殊符号的随机组合,与正常域名差异较大。基于时间的特征包括TTL和域名存活时间等,TTL值过短是DNS重绑定的一项重要特征。异常通信特征包括:域名地理位置、域名解析服务器及是否含有私有IP。恶意行为方面,在出现攻击时,域名服务器会对相同域名在短时间内进行多次解析。

文献[62]提出了一种两级检测的方案来判断DNS流量是否正常,第一级将反向DNS查找的CNAME主机名与原始域名进行比较,第二级比较DNS返回的IP地址的HTTP响应内容。文献[61]通过分析被动DNS数据,基于C4.5决策树、KNN、SVM及朴素贝叶斯等机器学习分类方法进行DNS重绑定检测。文献[63]提出一个实时恶意域名检测系统,使用CNN、LSTM深度学习算法自动提取特征,支持检测DNS重绑定相关域名。

3.2.7 恶意域名检测与分类技术

一些研究基于网络流量实现恶意域名检测与分类。Notos^[64]是一种基于被动DNS的域名信誉评价系统,该系统提取了网络、域名和黑名单证据3类特征,利用随机森林算法识别恶意软件传播、网络钓鱼和垃圾邮件活动。文献[65]提出的EXPOSURE系统,基于被动DNS数据,从时间、DNS响应、TTL值和域名4个维度选取了15个特征,利用J48决策树算法建立分类器,支持识别钓鱼网站、垃圾邮件、僵尸网络、Fast-Flux恶意域名等。文献[66]从DNS流量中提取了3组特征(词法特征、DNS统计特征和第三方特征),其中第三方特征主要是WHOIS信息和网页排名信息,通过SVM、KNN、MLP、高斯朴素贝叶斯(Gaussian naive Bayes, GNB)和对率回归等机器学习方法对恶意域名进行分类,能够识别垃圾邮件、网络钓鱼和恶意软件。

不同DNS异常检测使用的算法和方法如表2所示。

表2 不同DNS异常检测使用的算法和方法
Table 2 Different DNS anomaly detection algorithms and methods

DNS异常检测类别	文献	检测算法	检测方法
网络钓鱼检测	[25]	文本相似性度量	统计
	[26]	文本相似度量、视觉相似度量	统计
	[27]	SVM、朴素贝叶斯	机器学习
	[28]	SVM、KNN、决策树、随机森林	机器学习
	[29]	DBN	深度学习
域名不良应用检测	[30]	文本相似性度量	统计
	[31]	SVM、随机森林、对率回归、朴素贝叶斯、MLP	机器学习
	[32]	SVM、随机森林、对率回归	机器学习
	[33]	Bi-LSTM	深度学习
速变域名检测 (Fast-Flux)	[41]	加权SVM	机器学习
	[42]	层次聚类	机器学习
	[43]	MLP、RBF、SVM	机器学习
	[44]	多模态(GCN、Bi-LSTM、MLP、神经网络)	深度学习
速变域名检测 (Domain-Flux)	[45]	字符分布频率分析、域名加权分数	统计
	[46]	统计来自主机的NXDOMAIN请求	统计
	[47]	SVM	机器学习
	[48]	随机森林、X均值	机器学习
	[49]	LSTM和CNN	深度学习
	[50]	LSTM	深度学习
	[51]	CNN、LSTM、MLP结合安全多方计算和差分隐私两种隐私增强技术	深度学习
DNS隧道攻击检测	[56]	J48决策树	机器学习
	[57]	对率回归、SVM、决策树	机器学习
	[58]	会话行为、随机森林	机器学习
	[59]	CNN	深度学习
	[60]	DNN、CNN、RNN	深度学习
DNS重绑定攻击检测	[61]	C4.5决策树、KNN、SVM、朴素贝叶斯	机器学习
	[63]	CNN、LSTM	深度学习
恶意域名检测与分类	[64]	随机森林	机器学习
	[65]	J48决策树	机器学习
	[66]	SVM、KNN、MLP、高斯朴素贝叶斯、对率回归	机器学习

3.3 威胁情报管理技术

3.3.1 黑名单技术

黑名单技术可用于识别和阻止垃圾邮件、网络钓鱼和其他恶意网络活动。在应用中，通过检查邮件发送方的源IP地址、邮件服务器、网站URL等信息是否在这些黑名单中，来判断发件

人、网站或链接的可信度和安全威胁级别，从而过滤垃圾内容、阻止攻击。

一些黑名单对应威胁的类别，如垃圾邮件黑名单DNSBL、网络钓鱼黑名单PhishTank。DNSBL^[67]基于DNS协议来阻止接收来自黑名单列表中的邮件。PhishTank^[9]是允许用户提交、验

证和共享网络钓鱼链接的社区网站的, 并提供钓鱼网站检测接口。一些黑名单提供了一般性信息, 例如, 谷歌安全浏览 (Google safe browsing) 利用谷歌持续更新的恶意网址和钓鱼网站的数据集, 来检测URL的安全性。

在PDNS服务中使用的黑名单列表还包括: URLhaus、CyberCrime Tracker 和 VX Vault 等^[7]。URLhaus 共享有关用于恶意软件分发的恶意URL的情报; CyberCrime Tracker 提供了网络犯罪追踪情报信息; VX Vault 是恶意软件样本的情报信息。

3.3.2 CTI 技术

CTI^[68]技术是指有关潜在或实际网络威胁的信息, 这些信息包括威胁行为者所使用的策略、技术和程序信息, 以及危害指标, 如IP地址、域名和恶意软件的哈希值。CTI技术可用于通知决策者对网络威胁做出决策, 以避免受到攻击, 改善组织的安全状况。CTI技术的有效共享是网络威胁检测与预防的核心。由于CTI技术中的信息来自多种数据源, 需要进行标准化和结构化^[69]。

CTI技术信息的共享方法有^[70]: 结构化威胁信息表达式 (structured threat information expression, STIX)、指标信息的可信自动化交换 (trusted automated exchange of indicator information, TAXII) 和网络可观察表达式 (cyber observable expression, CybOX) 等结构化标准。STIX 是用于描述网络威胁信息的结构化语言, 它提供了基于标准XML的语法描述威胁情报的细节和威胁内容的方法。TAXII 是用于交换STIX数据的协议。每个STIX版本有对应的TAXII版本。CybOX 为STIX 1.x版本提供了表达网络可观测值、事件和其他属性的方法。

文献^[69]研究了CTI技术格式和语言的概况、公开的威胁信息源、实施方式, 分析了网络威胁情报源的样本、数据类型以及在聚合和共享数据时发现的问题。研究发现, CTI技术的数据来源与格式数量较多, 自定义的JSON格式是满足特定场景需求的快速而简单的方法。文献^[71]提出了THREATRAPTOR威胁搜寻系统, 可用于管理日志相关的网络威胁信息, 支持使用开源网络威胁情报提供的外部威胁信息。

3.4 数据存储管理技术

数据存储管理主要涉及数据仓库、数据湖以及湖仓一体 (lakehouse) 技术。数据仓库^[72]是一个面向主题的、集成的、非易失的、随时间变化的用来支持管理人员决策的数据集合。从形式上来说, 数据仓库是一个大型数据存储库, 可以以结构化的方式存储和集成各种来源的数据, 并通过适当的数据分析辅助决策^[73]。数据湖是集中式存储库, 用户能够以原始格式存储未处理的数据, 包括大规模的非结构化、半结构化或结构化数据。这些数据可以帮助企业通过大数据分析、机器学习和可视化做出更好的业务决策。湖仓一体^[74]是近年来出现的一种新的体系结构, 其目标是解决数据仓库和数据湖中存在的问题, 结合两者的优势, 它基于低成本和直接访问存储的数据管理系统, 还提供传统分析型数据库管理系统 (database management system, DBMS) 的功能。

SIDN的ENTRADA平台^[75]是基于Hadoop的开源高性能数据流仓库。CISA的PDNS服务使用数据湖来存储DNS递归解析服务器中DNS防护相关事件^[76], 进而支持对DNS安全威胁进行分析。

文献^[77]遵循湖仓一体架构模式, 提出了一个模块化数据平台的原型, DNS服务器的数据是平台的信息来源之一。该平台基于开放标准和开源组件, 遵循云原生原则, 可在公共云、私有云和混合云等现代计算环境中运行, 能够以可扩展的方式持续收集、处理、保留和分析大量网络测量数据。

4 发展趋势

4.1 问题和挑战

PDNS在发展过程中面临一些问题与挑战, 如威胁检测与处置导致的解析延迟、威胁情报的数据质量、集中化问题等。

(1) 解析延迟问题

文献^[7]对公共PDNS服务的部署和安全影响进行了大规模的测量研究。结果表明, 从客户端感知的角度, PDNS引起的查询延迟大多数情况下可以忽略不计。但PDNS对于时延的影响仍需进一步深入研究。比如, RPZ策略中包

含了大量域名与相应“动作”的映射关系，当策略数量过大、规则复杂时，DNS 服务器在查找对应规则时可能需要更长的时间，从而增加解析时延；如果配置了错误或不合理的 RPZ 规则，DNS 服务解析可能出错或陷入死循环，影响解析响应速度。

(2) 威胁情报的数据质量问题

PDNS 的防护效果依赖于威胁情报的准确性。黑名单列表存在因数据源或使用用户而导致的更新不及时的风险。威胁情报多源异构，质量和可信度参差不齐，这也给威胁情报的存储、共享及应用带来了难题。文献[78]定义了一组威胁情报数据源的指标，用于系统地表征公共和商业来源，并对 CTI 信息进行定量评估，结果表明，CTI 数据存在限制和挑战，如一些 CTI 中大量数据没有明确的类别定义；不同类别、不同指标的数据源数据，质量存在明显差异。

(3) 集中化问题

随着 PDNS 在安全方面发挥作用的提升，用户的 DNS 查询会产生集中化趋势^[6]，这会带来一些潜在的安全风险和集中化问题。如果大量用户的 DNS 查询都依赖于同一个中心化服务，一旦该服务出现故障或遭受攻击，将导致用户的网络应用受到影响。此外，集中化的 PDNS 服务会涉及数据收集与分析，一旦遭到黑客攻击，会存在数据泄露、DNS 劫持等风险。

4.2 趋势和研究方向

(1) 新型 DNS 相关安全威胁的检测

针对 DNS 新型安全问题，一些传统方法的应用效果将受到限制或不再适用。例如，在利用 DNS 信道传输加密协议的 DNS 隧道攻击中，检测模型难以从 DoH 加密后的 DNS 流量中获取 DNS 请求响应数据包的载荷，一些基于签名和基于阈值的方法不再适用。目前对 DoH 恶意流量的检测研究主要使用机器学习算法^[79]，更为高效和智能的检测方法还需进一步研究。

(2) 新技术在 DNS 异常检测中的应用

Transformer、BERT (bidirectional encoder representations from transformer) 等技术在 DNS 异常检测中应用前景广阔。在针对日志类的大规模文本数据分析场景中，一般使用深度学习

RNN 算法。与传统的 RNN 相比，Transformer 模型具有更好的并行计算性能、更好的长程依赖建模能力。BERT 是基于 Transformer 架构的预训练语言模型，它通过在大规模文本数据上的预训练来捕捉语言的深层双向表征。在实际任务中，使用具体任务数据进行训练、微调，以取得良好的效果。目前已有研究基于 BERT 对 DNS 日志数据进行处理分析，如文献[80]使用 BERT 来学习网络中的完全限定域名，进行上下文嵌入，以提高深度模型的泛化性能。

(3) DNS 威胁情报信息的动态管理

威胁情报管理需要覆盖情报的整个生命周期，从情报的收集、分析、分发、应用到废弃，确保情报的及时更新和准确性。随着网络攻击的不断演变，实时共享最新的威胁情报变得至关重要。网络安全用户更加依赖于动态的情报共享平台，以便快速获取和分发最新的威胁信息。

(4) PDNS 与其他网络安全平台的协同联动

不同网络安全实体或系统之间通过信息共享、协调行动和资源整合，可以提高整体网络安全防护能力。PDNS 需要获取外部威胁情报信息源推送的最新威胁情报信息，以便及时应用到 DNS 处理策略中，提高防御效率。另外，及时将 PDNS 检测、分析得到的威胁情报信息共享给外部网络安全防御平台，有助于网络安全平台针对新型安全威胁做出快速响应。

5 结束语

本文介绍了 PDNS 的概念、功能和目前的应用现状，对 PDNS 所涉及的关键技术如域名威胁处置技术、DNS 异常检测技术、威胁情报管理技术和数据存储管理技术等，进行了综述研究。最后分析了 PDNS 目前面临的问题与挑战，并对 PDNS 未来发展趋势与研究方向提出了展望。

参考文献：

- [1] FOUCHEREAU B. 2023 Global DNS threat report[R]. IDC. 2023.
- [2] DNS security threat mitigation program[EB]. ICANN. 2023.
- [3] Selecting a protective DNS service[EB].CISA. 2021.
- [4] RADU R. DNS4EU: a step change in the EU's strategic autonomy?[J]. Journal of Cyber Policy, 2023, 8(2): 239-256.
- [5] CARR A, ALAM A, ALLISON J. Monitoring malicious DNS queries: an experimental case study of utilising the national cyber secu-

- city centre's protective DNS within a UK public sector organisation [EB]. 2023.
- [6] RODRIGUEZ E, ANGHEL R, PARKIN S, et al. Two sides of the shield: understanding protective DNS adoption factors[C]//Proceedings of 32nd USENIX Security Symposium, 2023: 3135-3152.
- [7] LIU M X, ZHANG Y M, LI X, et al. Understanding the implementation and security implications of protective DNS services[C]//Proceedings of 2024 Network and Distributed System Security Symposium. Reston, VA: Internet Society, 2024.
- [8] MOCKAPETRIS P. Domain names-concepts and facilities[R]. 1987.
- [9] 张茜, 延志伟, 李洪涛, 等. 网络钓鱼欺诈检测技术研究[J]. 网络与信息安全学报, 2017, 3(7): 7-24.
- ZHANG X, YAN Z W, LI H T, et al. Research of phishing detection technology[J]. Chinese Journal of Network and Information Security, 2017, 3(7): 7-24.
- [10] 临江. 去年全国受理网络违法和不良信息举报 1.72 亿件[N], 人民邮电, 2023, 2: 10.
- LIN J. Last year, 172 million reports of illegal and harmful information on the Internet were received nationwide[N]. People's Posts and Telecommunications, 2023, 2: 10.
- [11] 章思宇. 基于DNS流量的恶意软件域名挖掘[D]. 上海: 上海交通大学, 2014.
- ZHANG S Y. Detecting malware domains on DNS traffic[D]. Shanghai: Shanghai Jiao Tong University, 2014.
- [12] VORMAYR G, ZSEBY T, FABINI J. Botnet communication patterns[J]. IEEE Communications Surveys & Tutorials, 2017, 19(4): 2768-2796.
- [13] JEITNER P, SHULMAN H. Injection attacks reloaded: tunnelling malicious payloads over DNS[C]// Proceedings of 30th USENIX Security Symposium (USENIX Security 21). 2021: 3165-3182.
- [14] JACKSON C, BARTH A, BORTZ A, et al. Protecting browsers from DNS rebinding attacks[C]// Proceedings of the 14th ACM Conference on Computer and Communications Security. New York: ACM, 2007: 421-431.
- [15] ACAR G, HUANG D Y, LI F, et al. Web-based attacks to discover and control local IoT devices[C]// Proceedings of the 2018 Workshop on IoT Security and Privacy. New York: ACM, 2018: 29-35.
- [16] JENSEN B, NAIR D, ATALAN Y, et al. CISA's evolving .gov mission: defending the United States' federal executive agency networks [R]. 2023.
- [17] European Commission. The EU's cybersecurity strategy for the digital decade[EB]. 2020.
- [18] HESSELMAN C, MOURA G, SCHMIDT R, et al. Increasing DNS security and stability through a control plane for top-level domain operators[J]. IEEE Communications Magazine, 2017, 55: 197-203.
- [19] VIXIE P, SCHRYVER V. DNS response policy zones (RPZ)[S]. Internet Engineering Task Force 2018.
- [20] WILDE N, JONES L, LOPEZ R, et al. A DNS RPZ firewall and current American DNS practice[C]// Proceedings of International Conference on Information Science and Applications. Singapore: Springer, 2019: 259-265.
- [21] ICHISE H, JIN Y, IIDA K. Policy-based detection and blocking system for abnormal direct outbound DNS queries using RPZ[C]// Proceedings of 2022 the 12th International Workshop on Computer Science and Engineering. WCSE, 2022: 327-332.
- [22] KIM H, CHOI S S, SONG J. A methodology for multipurpose DNS sinkhole analyzing double bounce emails[C]// Proceedings of Neural Information Processing. Berlin Heidelberg: Springer 2013: 609-616.
- [23] LI Y C, CHU S N, XIAO R. A pharming attack hybrid detection model based on IP addresses and web content[J]. Optik, 2015, 126 (2): 234-239.
- [24] CHIEW K L, CHANG E H, SZE S N, et al. Utilisation of website logo for phishing detection[J]. Computers & Security, 2015, 54: 16-26.
- [25] 邹学强, 张鹏, 黄彩云, 等. 基于页面布局相似性的钓鱼网页发现方法[J]. 通信学报, 2016, 37(S1): 116-124.
- ZOU X Q, ZHANG P, HUANG C Y, et al. Phishing attacks discovery based on HTML layout similarity[J]. Journal on Communications, 2016, 37(S1): 116-124.
- [26] RAO R S, PAIS A R. Jail-Phish: an improved search engine based phishing detection system[J]. Computers and Security, 2019, 83(C): 246-267.
- [27] JAIN A K, GUPTA B B. PHISH-SAFE: URL features-based phishing detection system using machine learning[C]// Proceedings of Cyber Security. Singapore: Springer, 2018: 467-474.
- [28] PURBAY M, KUMAR D. Split behavior of supervised machine learning algorithms for phishing URL detection[C]// Proceedings of Advances in VLSI, Communication, and Signal Processing. Singapore: Springer, 2021: 497-505.
- [29] YI P, GUAN Y X, ZOU F T, et al. Web phishing detection using a deep learning framework[J]. Wireless Communications and Mobile Computing, 2018, 2018(1): 4678746.
- [30] 尚秋明, 王利军, 邓桂英, 等. 一种不良域名快速核验方法的研究[J]. 电子技术应用, 2022, 48(10): 72-77.
- SHANG Q M, WANG L J, DENG G Y, et al. Research on a fast verification method for malicious domain names[J]. Application of Electronic Technique, 2022, 48(10): 72-77.
- [31] MIN M, LEE J, LEE K. Detecting illegal online gambling (IOG) services in the mobile environment[J]. Security and Communication Networks, 2022 (1): 3286623.
- [32] CHEN Y, ZHENG R, ZHOU A, et al. Automatic detection of pornographic and gambling websites based on visual and textual content using a decision mechanism[J]. Sensors, 2020, 20(14): 3989.
- [33] WANG C Y, ZHANG M, SHI F, et al. A hybrid multimodal data fusion-based method for identifying gambling websites[J]. Electronics, 2022, 11(16): 2489.
- [34] GRAHAM M, WINCKLES A, MOORE A. Botnet detection in virtual environments using Netflow[C]// Proceedings of 7th International Conference on Cybercrime Forensics Education and Training, UK. 2014.
- [35] GU G, PORRAS P A, YEGNESWARAN V, et al. BotHunter: detecting malware infection through IDS-driven dialog correlation [C]// Proceedings of 16th USENIX Security Symposium. USENIX Association, 2007.
- [36] RAHBARINIA B, PERDISCI R, ANTONAKAKIS M. Efficient and accurate behavior-based tracking of malware-control domains in large ISP networks[J]. ACM Transactions on Privacy and Security

- rity, 2016, 19(2): 4.
- [37] CHATZIS N, POPESCU-ZELETIN R. Flow level data mining of DNS query streams for email worm detection[C]// Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS' 08. Berlin Heidelberg: Springer, 2009: 186-194.
- [38] FERNANDEZ S, KORCZYNSKI M, DUDA A. Early detection of spam domains with passive DNS and SPF[C]// Proceedings of International Conference on Passive and Active Network Measurement. Cham: Springer, 2022: 30-49.
- [39] CHOI H, LEE H, LEE H, et al. Botnet detection by monitoring group activities in DNS traffic[C]// Proceedings of the 7th IEEE International Conference on Computer and Information Technology (CIT 2007). Piscataway: IEEE Press, 2007: 715-720.
- [40] YADAV S, NARASIMHA REDDY A L. Winning with DNS failures: strategies for faster botnet detection[C]// Proceedings of International Conference on Security and Privacy in Communication Systems. Berlin Heidelberg: Springer, 2012: 446-459.
- [41] YU X Z, ZHANG, B, KANG L, et al. Fast-Flux botnet detection based on weighted SVM [J]. Information Technology Journal, 11 (8): 1048-1055.
- [42] PERDISCI R, CORONA I, GIACINTO G. Early detection of malicious flux networks via large-scale passive DNS traffic analysis[J]. IEEE Transactions on Dependable and Secure Computing, 2012, 9 (5): 714-726.
- [43] AL-DUWAIIRI B, JARRAH M, SHATNAWI A. PASSVM: a highly accurate fast flux detection system[J]. Computers & Security, 2021.110: 102431.
- [44] 郎波, 谢冲, 陈少杰, 等. 基于多模式特征融合的Fast-Flux恶意域名检测方法[J]. 信息安全学报, 2022, 22(4): 20-29.
- LANG B, XIE C, CHEN S J, et al. Fast-flux malicious domain Name detection method based on multimodal feature fusion[J]. Netinfo Security, 2022, 22(4): 20-29.
- [45] AGYEPONG E, BUCHANAN W J, JONES K. Detection of algorithmically generated malicious domain[C]//Proceedings of the Computer Science & Information Technology. Academy & Industry Research Collaboration Center (AIRCC). 2018: 13-32.
- [46] PATSAKIS C, CASINO F. Exploiting statistical and structural features for the detection of domain generation algorithms[J]. Journal of Information Security and Applications, 2021, 58: 102725.
- [47] WANG Z, JIA Z T, ZHANG B. A detection scheme for DGA domain names based on SVM[C]//Proceedings of the 2018 International Conference on Mathematics, Modelling, Simulation and Algorithms (MMSA 2018). Paris, France: Atlantis Press, 2018: 257-263.
- [48] 于光喜, 张棣, 崔华俊, 等. 基于机器学习的僵尸网络DGA域名检测系统设计与实现[J]. 信息安全学报, 2020, 5(3): 35-47.
- YU G X, ZHANG Y, CUI H J, et al. Design and implementation of a DGA domain Name detection system by machine learning[J]. Journal of Cyber Security, 2020, 5(3): 35-47.
- [49] YU B, GRAY D L, PAN J, et al. Inline DGA detection with deep networks[C]// Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW). Piscataway: IEEE Press, 2017: 683-692.
- [50] YILMAZ I, SIRAJ A, ULYBYSHEV D. Improving DGA-based malicious domain classifiers for malware defense with adversarial machine learning[C]// Proceedings of the 2020 IEEE 4th Conference on Information & Communication Technology (CICT). Piscataway: IEEE Press, 2020: 1-6.
- [51] MAIA R J M, RAY D, PENTYALA S, et al. An end-to-end framework for private DGA detection as a service[J]. Cryptology ePrint Archive, 2023.
- [52] SAMMOUR M, HUSSIN B, OTHMAN M F I, et al. DNS tunneling: a review on features[J]. International Journal of Engineering & Technology, 2018, 7(3.20): 1.
- [53] WANG Y, ZHOU A M, LIAO S, et al. A comprehensive survey on DNS tunnel detection[J]. Computer Networks, 2021, 197: 108322.
- [54] SHERIDAN S, KEANE A. Detection of DNS-based covert channel beacon signals[J]. Journal of Information Warfare, 2015, 14(4): 100-114.
- [55] BORN K, GUSTAFSON D. Detecting DNS tunnels using character frequency analysis[J]. arXiv Preprint arXiv: 1004.4358, 2010.
- [56] 章思宇, 邹福泰, 王鲁华, 等. 基于DNS的隐蔽通道流量检测[J]. 通信学报, 2013, 34(5): 143-151.
- ZHANG S Y, ZOU F T, WANG L H, et al. Detecting DNS-based covert channel on live traffic[J]. Journal on Communications, 2013, 34(5): 143-151.
- [57] LIU J K, LI S H, ZHANG Y Z, et al. Detecting DNS tunnel through binary-classification based on behavior features[C]// Proceedings of the 2017 IEEE Trustcom/BigDataSE/ICSS. Piscataway: IEEE Press, 2017: 339-346.
- [58] ZHAO Y, YE H Z, LI L Z, et al. Detecting DNS tunnels using session behavior and random forest method[C]// Proceedings of the 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC). Piscataway: IEEE Press, 2020: 45-52.
- [59] LIU C, DAI L, CUI W J, et al. A byte-level CNN method to detect DNS tunnels[C]// Proceedings of the 2019 IEEE 38th International Performance Computing and Communications Conference (IPCCC). Piscataway: IEEE Press, 2019: 1-8.
- [60] ZHANG J C, YANG L, YU S, et al. A DNS tunneling detection method based on deep learning models to prevent data exfiltration [C]// Proceedings of International Conference on Network and System Security. Cham: Springer, 2019: 520-535.
- [61] 郭恒臻, 潘祖烈, 沈毅, 等. 一种基于被动DNS数据分析的DNS重绑定攻击检测技术[J]. 信息安全学报, 2021, 21(3): 87-95.
- GUO X Z, PAN Z L, SHEN Y, et al. DNS rebinding detection technology based on passive DNS data analysis[J]. Netinfo Security, 2021, 21(3): 87-95.
- [62] BRAHMASANI S, SIVASANKAR E. Two level verification for detection of DNS rebinding attacks[J]. International Journal of System Assurance Engineering and Management, 2013, 4(2): 138-145.
- [63] SPAULDING J, MOHAISEN A. Defending Internet of things against malicious domain names using D-FENS[C]//Proceedings of the 2018 IEEE/ACM Symposium on Edge Computing (SEC). Piscataway: IEEE Press, 2018: 387-392.
- [64] ANTONAKAKIS M, PERDISCI R, DAGON D, et al. Building a dynamic reputation system for DNS[C]// Proceedings of 19th USENIX Security Symposium, Washington: USENIX Association,

- 2010.
- [65] BILGE L, KIRDA E, KRUEGEL C, et al. EXPOSURE: finding Malicious domains using passive DNS analysis[C]// Proceedings of the Network and Distributed System Security Symposium (NDSS). 2011: 1-17.
- [66] MAHDAVIFAR S, MALEKI N, LASHKARI A H, et al. Classifying malicious domains using DNS traffic analysis[C]// Proceedings of the 2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/Cyber-SciTech). Piscataway: IEEE Press, 2021: 60-67.
- [67] JUNG J, SIT E. An empirical study of Spam traffic and the use of DNS black lists[C]// Proceedings of the 4th ACM SIGCOMM conference on Internet measurement. New York: ACM, 2004: 370-375.
- [68] MAGNUSSON J. Survey and analysis of DNS filtering components[EB]. arXiv preprint arXiv: 2401.03864, 2024.
- [69] RAMSDALE A, SHIAELES S, KOLOKOTRONIS N. A comparative analysis of cyber-threat intelligence sources, formats and languages[J]. Electronics, 2020, 9(5): 824.
- [70] TSOICHEV G, YOSHINOV R. The research on intelligent DNS security[C]//Proceedings of the 2020 IEEE 10th International Conference on Intelligent Systems (IS). Piscataway: IEEE Press, 2020: 409-414.
- [71] 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.
- [72] INMON W H. 数据仓库[M]. 北京: 机械工业出版社, 2006. INMON W H. Data warehouse[M]. Beijing: China Machine Press, 2006.
- [73] CHANDRA P, GUPTA M K. Comprehensive survey on data warehousing research[J]. International Journal of Information Technology, 2018, 10(2): 217-224.
- [74] ARMBRUST M, GHODSI A, XIN R, et al. Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics[C]//Proceedings of CIDR. 2021, 8: 28.
- [75] WULLINK M, MOURA G C M, MÜLLER M, et al. ENTRADA: a high-performance network traffic data streaming warehouse[C]// Proceedings of the NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium. Piscataway: IEEE Press, 2016: 913-918.
- [76] Protective domain name system resolver service[EB]. CISA. 2023.
- [77] TOVARŇÁK D, RAČEK M, VELAN P. Cloud native data platform for network telemetry and analytics[C]// Proceedings of the 2021 17th International Conference on Network and Service Management (CNSM). Piscataway: IEEE Press, 2021: 394-396.
- [78] LI V G, DUNN M, PEARCE P, et al. Reading the tea leaves: a comparative analysis of threat intelligence. [C]// Proceedings of 28th USENIX Association, 2019: 851-867.
- [79] DING S, ZHANG D Q, GE J G, et al. Encrypt DNS traffic: automated feature learning method for detecting DNS tunnels[C]// Proceedings of the 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking. Piscataway: IEEE Press, 2021: 352-359.
- [80] LE F, WERTHEIMER D, CALO S, et al. NorBERT: Network representations through BERT for network analysis & management[C]// Proceedings of the 2022 30th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS). Piscataway: IEEE Press, 2022: 25-32.

[作者简介]



马永征 (1977-), 男, 山西运城人, 中国互联网络信息中心正高级工程师, 主要研究方向为分布式计算、图计算、下一代互联网。



李洪涛 (1977-), 男, 河北保定人, 中国互联网络信息中心正高级工程师, 主要研究方向为互联网基础资源大数据、网络安全、下一代互联网。



马中胜 (1979-), 男, 吉林长春人, 中国互联网络信息中心工程师, 主要研究方向为网络安全、下一代互联网。



胡卫宏 (1979-), 男, 陕西韩城人, 中国互联网络信息中心高级工程师, 主要研究方向为网络通信与网络信息安全。



张中献 (1995-), 男, 北京平谷人, 中国互联网络信息中心助理工程师, 主要研究方向为图计算、神经网络模型、参数估计、下一代互联网。