

基于动态域名水印的IPv6 DNS服务发现方法

韩丁康^{1,2}, 朱宇佳^{1,2}, 赵蕾^{1,2}, 焦亮^{1,2}, 刘庆云^{1,2}

(1.中国科学院信息工程研究所, 北京 100085; 2.中国科学院大学网络空间安全学院, 北京 100049)

摘要: IPv6环境下的域名系统(DNS, domain name system)服务发展迅速,开展IPv6环境下DNS服务发现方法研究,对分析DNS服务在IPv6环境下的全球发展态势,提升服务可靠性具有非常重要的意义。基于IPv4和IPv6的合作解析关系,通过跨栈服务关联发现IPv6 DNS服务是一种有效的方法。然而,现有基于跨栈服务关联的IPv6 DNS服务发现方法受DNS探测包长度限制,探测能力有限。针对此问题,提出一种基于动态域名水印的IPv6 DNS服务发现方法。该方法利用自建权威服务器构建动态域名资源记录,绕过探测包长度限制。相比传统方法,该方法发现的IPv6 DNS服务数量提升接近98%;同时,通过解析水印日志记录,发现了解析器间存在大量解析依赖和集中化现象。

关键词: 域名系统; IPv6; 服务测绘; 域名水印

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.2096-109x.2024066

Method for the discovery of IPv6 DNS services based on the dynamic domain name watermarks

HAN Dingkang^{1,2}, ZHU Yujia^{1,2}, ZHAO Lei^{1,2}, JIAO Liang^{1,2}, LIU Qingyun^{1,2}

1. Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100085, China

2. School of Cyberspace Security, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract: DNS (domain name system) services in the IPv6 environment experienced rapid development. Research on methods for discovering DNS services in an IPv6 environment was of paramount importance in order to analyze the global developmental trends of DNS services in IPv6 and enhance their reliability. Based on the cooperative resolution relationship between IPv4 and IPv6, associating IPv6 DNS services through inter-stack services was an effective approach. However, existing methods for discovering IPv6 DNS services based on inter-stack service association were limited by the length of DNS probe packets. In response to this issue, a novel approach was proposed for discovering IPv6 DNS services based on dynamic domain name watermarks. By constructing dynamic domain name resource records using self-established authoritative servers, the limitations imposed by probe packet length were circumvented. Comparative analysis indicated that the number of discovered IPv6 DNS services using this method increased by nearly 98% compared to traditional methods. Moreover, through the analysis of watermark

收稿日期: 2024-05-27; 修回日期: 2024-07-22

通信作者: 朱宇佳, zhuyujia@iie.ac.cn

引用格式: 韩丁康, 朱宇佳, 赵蕾, 等. 基于动态域名水印的IPv6 DNS服务发现方法[J]. 网络与信息安全学报, 2024, 10(5): 56-70.

Citation Format: HAN D K, ZHU Y Y, ZHAO L, et al. Method for the discovery of IPv6 DNS services based on the dynamic domain name watermarks[J]. Chinese Journal of Network and Information Security, 2024, 10(5): 56-70.

logs, a plethora of resolver dependencies and centralization phenomena were identified.

Keywords: domain name system, IPv6, service measurement, domain name watermark

0 引言

域名系统（DNS，domain name system）是重要的互联网基础设施之一，承担着域名和IP地址间相互转换的任务，被誉为Web世界的“指南针”。自DNS诞生之日起，DNS信息的传递一直依赖于网络层的IPv4。然而，随着互联网的快速发展，对IP地址的需求日益增长，32位长度的IPv4地址难以应对。2020年，AFRINIC（非洲网络信息中心）完成了最后一个IPv4可用地址池的分配，宣告全球所有的（约43亿个）IPv4地址全部分配完毕。2023年，亚马逊宣布于2024年2月1日开始对其管辖下所有公共IPv4地址进行收费。这无疑增加了IPv4应用服务的运维成本。IPv4地址的短缺限制了DNS服务的进一步发展，亟须新的解决措施。

为了应对IPv4存在的地址空间不足的问题，IPv6于1993年被提出。IPv6具有128位的地址长度，拥有“可以为地球上每一粒沙子分配一个地址”的庞大地址空间。大力发展IPv6已成为互联网升级演进的必然趋势，是我国实现网络强国建设的战略基础支撑。IPv6与DNS的发展前景是紧密相联的。一方面，IPv6在发展中的广泛应用离不开作为互联网基础设施的DNS服务的有力支持；另一方面，IPv6海量的地址空间扩展了DNS服务进一步发展的可能性。目前，主流的操作系统和浏览器客户端均已实现基于IPv6的DNS查询支持；以Bind9为代表的DNS服务组件也实现了在IPv6环境下域名解析服务的适配；谷歌、阿里云和DNSPod等大型公共DNS服务商先后开放了提供公共DNS服务的IPv6入口地址。IPv6环境下的DNS服务部署正呈现迅猛发展之势。

从国家网络安全治理的角度来看，IPv6的快速发展增加了对其进行服务发现和安全监测的必要性。这需要尽可能地发现全网已部署的IPv6 DNS服务^①，绘制网络资源地图。然而，IPv6地址空间巨大，使用常规探测技术单轮次探测所需

时间是IPv4的 2^96 倍，需要付出的时间和计算成本难以承受，相关的安全监测能力尚且处于起步阶段。因此，我国已将构建IPv6安全监测体系，提高IPv6安全态势感知能力作为“十四五”时期强化国家网络安全保障能力措施中的重要一环。

为了解决IPv6环境下DNS服务测绘技术面临的上述难题，本文提出一种基于动态水印的IPv6 DNS服务发现方法，该方法的创新之处在于使用动态域名水印实现从单次探测中发现、收集更多的IPv6 DNS服务地址，提高IPv6 DNS服务的挖掘效率。

1 相关工作

在IPv4环境下，常通过对全IPv4地址空间（ 2^{32} ）发送DNS请求报文的方式发现DNS服务节点。而IPv6地址由一个长度为128位的字符串组成，拥有 2^{128} 的寻址空间，其结构如图1所示。

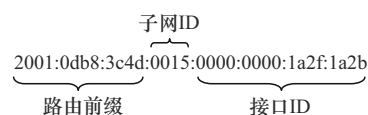


图1 IPv6地址的结构

Figure 1 IPv6 addresses structure

IPv6的地址空间远大于IPv4的地址空间，通过暴力扫描全地址空间发现IPv6中的DNS服务几乎不可能，因此需要新的高效的服务发现手段。为了解决这一难题，学者提出了多种高效发现IPv6环境中DNS服务的方法。

当前IPv6 DNS服务发现方法主要包括基于种子地址的主动发现技术、基于跨栈解析服务关联的主动发现技术以及基于数据收集的被动方法3类，如表1所示。其中，基于数据收集的被动方法需要部署网络监控节点或者用户侧脚本来收集IPv6下的DNS请求和响应流量，对部署节点和脚本有较高的实验条件要求，难以轻易达成。因此，本文主要介绍基于种子地址的主动发现技术和基于跨栈解析服务关联的主动发现技术。

^①本文中“IPv6 DNS服务”指部署在IPv6环境下且基于IPv6传输信息的DNS解析服务。

表 1 相关工作分类
Table 1 Classification of related works

方法类型	方法思想	方法
主动探测	基于种子地址预测	基于深度学习 ^[1-2] 基于地址空间树 ^[3-4]
	基于跨栈解析服务 关联发现	基于 IPv4 开放解析器 ^[5-6]
被动收集	被动的数据收集	部署网络监控节点 ^[7]

1.1 基于种子地址的主动发现技术

基于种子地址的主动发现方法假设存在活跃 IPv6 地址集中分布的区域。其技术思想是：首先，收集一系列活跃 IPv6 地址作为种子地址数据集，通过学习这些地址的内在分布特征和模式，预测生成可能存活的 IPv6 地址范围；接下来，借助扫描工具收集这一范围内实际活跃的 IPv6 地址，并合入种子地址集合中，形成迭代预测的效果。通过向已知的活跃地址发送 DNS 请求，验证其是否搭载 DNS 服务。根据学习模型的不同，可以分为基于深度学习的方法和基于地址空间树的方法。

基于深度学习的方法将 IPv6 地址视为一个存在语义关系的字符串。Cui 等^[1]通过将 IPv6 地址转化为空间向量，使用 Transformer 模型对向量内的语义关系进行挖掘与学习。利用 IPv6 地址的 64 位前缀对后 64 位接口 ID 进行预测，完成 IPv6 地址的预测生成。无独有偶，基于生成对抗网络（GAN, generative adversarial networks）强大的预测生成能力，6GAN^[2]首先将 IPv6 地址进行模式分类，然后针对每一类别利用 GAN 进行特征学习与生成，从而根据每一类地址模式预测可能活跃的地址。这些方法必须建立庞大的深度学习模型，对算力要求较高，应对大规模的网络探测需要更高的成本。

基于地址空间树的方法将 IPv6 地址理解为高维度的向量，对种子地址进行聚类或者分裂聚类，生成被称为空间树的数据结构，该结构中的每一个节点表征了不同维度的值的变化，通过优先扫描空间树上的高密度区域，可以获取更多的可能活跃地址。Liu 等^[3]首次引入该数据结构并提出了基于该思想的 6Tree。目前基于地址空间树的方法面临离群地址（outlier address）带来的挑战。这些地址会影响地址空间树的生成，使得节点对应

的地址范围指数级增加，降低了活跃地址的发现效率。为了解决这一问题，Yang 等^[4]利用图模型对生成的空间树进行建模，检测种子地址中的离群地址，并对原始地址空间树进行枝的裁剪，以减少对低效区域的探测，从而提高模型的预测效果。

基于种子地址的主动发现技术虽然取得了显著的成效，但其有效性在很大程度上受限于初始种子地址集的多样性和覆盖范围。近年来，为了隐蔽递归解析器的存在，降低被扫描发现的概率，大部分运营商为递归解析器配置了严格的访问控制策略，因此已有的 IPv6 DNS 种子地址集中大部分是开放解析器，而递归解析器的数量仅有不足 1 000 台。基于此类方法很难发现更多递归解析器。

1.2 基于跨栈解析服务关联的主动发现技术

1.2.1 跨栈域名解析

DNS 解析器在进行域名记录查询时，会从根服务器开始查询，接下来进行迭代查询，直到找到该域名对应的权威服务器。

随着近年来 IPv6 网络的快速发展，越来越多的 DNS 服务组件被部署到 IPv6 网络中。然而，IPv6 和 IPv4 主机不能直接通信，解析器无法从工作在另一个版本的 IP 协议栈上的权威服务器获取记录，导致 DNS 的正常运行受到影响。

为了解决跨栈问题对域名解析过程的干扰，一些研究人员提出对现有的 DNS 体系进行改造，使 IPv4 和 IPv6 DNS 服务之间可以合作完成域名解析。这种合作关系可以分为基于转发和基于双栈两类，两类合作解析形式示意（以 IPv4 解析器与 IPv6 权威服务器为例）如图 2 所示。

（1）基于转发的合作关系

基于转发的合作解析方案的思路是将无法处理的跨栈解析过程委托给其他可处理的 DNS 服务器进行实际的解析，并等待该服务器返回最终结果，最终完成域名解析过程。研究人员根据支持协议的不同为解析器设计了不同的委托方案。对于 IPv4 DNS 解析器，可以将请求直接转发到上游的双栈解析器的 IPv4 地址入口，利用该服务器或其他 IPv6 DNS 解析器的 IPv6 通信能力获取资源记录^[6]；对于 IPv6 DNS 解析器，则可以利用 DNS64+NAT64 的组合方案^[8]，实现 DNS 的请求和响应在不同协议之间的转换；特别地，由于 DNS64+NAT64 实际充当了类似

于双栈服务器的角色, 本文将其归类为基于转发的合作解析方案的一种实现方式。基于转发的合作解析实现简单, 且对于现有网络改造侵入力度小, 因而广受欢迎。

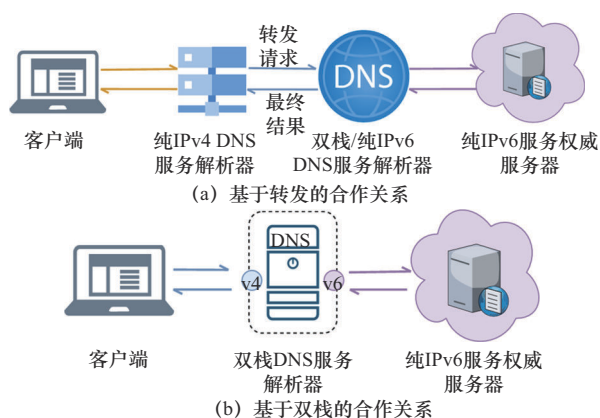


图2 两类合作解析形式示意

(以IPv4解析器与IPv6权威服务器为例)

Figure 2 Representation of two distinct forms of collaborative resolution (taking the example of an IPv4 resolver and an IPv6 authoritative server)

(2) 基于双栈的合作关系

基于双栈的合作解析方案该方案的核心思想是对DNS服务的网络环境进行改造, 以满足从IPv4到IPv6的过渡需求。通过将现有的单栈IPv4或IPv6 DNS服务直接改造为双栈服务, DNS服务将同时具备在两种不同协议上提供服务的能力, 是最为理想的改造方案。然而, 将DNS服务从单栈部署改造为双栈部署需要对底层网络环境进行升级, 同时增加针对IPv6的各类辅助设施, 包括额外的防火墙设置、新的路由规则等, 对运营者的管理能力要求较高。

1.2.2 相关研究

基于此, 一些研究人员将目光放到了存在已久的开放IPv4 DNS服务上。2013年, Berger等^[5]提出了一种基于IPv4 DNS服务关联和IP信息水印嵌入的主动IPv6服务发现技术。他们利用跨栈域名解析的解析过程, 在可控域下配置一组相互独立工作的IPv4-only和IPv6-only域名服务器, 分别负责一个子域名的解析。两个子域名被配置互为对方的CNAME指向, 并且权威服务器将IP信息追加嵌入域名水印中。该方法只需向IPv4解析器发送探针, 就可以在客户端收集到存在解析

关联的IPv6解析器地址。Al-Dalky等^[9]还通过在CDN服务商处收集DNS请求记录的方法来获取IPv6解析器的地址。对来自同一子网的用户请求进行汇聚, 提取出该网络中对应的IPv4和IPv6解析器, 从而完成对IPv6域名解析服务的挖掘。这类方法注重收集与权威服务器进行通信的IPv6 DNS递归解析器, 因而可以在探测活跃IPv6地址的同时, 确认该地址是否存在DNS服务。

1.3 研究现状总结

众多学者围绕IPv6 DNS服务测绘技术开展了大量研究, 特别是活跃服务的分布测量, 并取得了阶段性进展, 但仍有许多关键问题亟待解决。首先, 目前关于IPv6 DNS服务发现的研究方法主要发现的是可公开访问的开放转发器, 对关键的直接与权威服务器建立通信的IPv6 DNS递归解析器缺乏更多的研究支持。这些解析器往往使用了严格的访问控制手段, 难以直接探测发现。现有的可挖掘递归IPv6 DNS服务方法由于采用静态域名水印, 不断增加的水印长度将会触发DNS协议标准中对域名长度的限制, 探测效率受限。因此, 本文将注重研究更高效的递归IPv6 DNS服务的发现方法。

2 本文方法

从DNS的服务架构来看, 解析器可以分为转发和递归两类。转发器将请求转发至上游解析器并等待返回响应, 而递归解析器则直接向权威服务器进行查询, 从源头决定着域名记录的正确性。因而, 绘制递归解析器的网络资源地图具有更重要的意义。近年来, 为了隐蔽递归解析器的存在, 降低被扫描发现的概率, 大部分运营商为递归解析器配置了严格的访问控制策略, 因而大部分递归解析器并不能通过客户端直接发送请求的方式发现。

目前, 围绕着IPv6递归解析器的探测发现, Berger^[5]等提出了基于静态域名水印的方法, 利用向开放IPv4 DNS解析器发送特殊请求以挖掘IPv6递归解析器。作为DNS研究中的重要探测手段^[10]之一, 域名水印由特定形式的子域名组成, 形成唯一标识, 对于DNS体系结构等相关研究具有重要作用。Al-Dalky等^[9]通过向开放的IPv4 DNS解析器发送查询请求, 触发IPv6解析器参与到精心设计的跨栈解析过程中, 从而在自

建权威服务器处将 IPv6 地址编码为水印嵌入域名中, 在客户端得到响应结果并解析水印, 还原 IPv4 与 IPv6 DNS 服务间关联关系, 最终发现 IPv6 DNS 服务。

前人采用的静态域名水印存在两个弊端。一是受域名长度的制约, 限制了探测效率。二是前人方法将信息嵌入域名中, 并在客户端进行关联关系的还原, 难以处理目前常见的多出口递归解析器的 DNS 服务架构, 造成数据分析不全面。因而, 本文提出一种基于动态域名水印的 IPv6 DNS 服务发现方法, 通过构建动态域名资源记录嵌入水印, 在自建 DNS 权威服务器处发现 IPv6 DNS 服务, 基于动态域名水印的 IPv6 DNS 服务发现方法的架构如图 3 所示。

2.1 跨栈域名解析场景构建

本文方法的核心原理是向 IPv4 解析器发送特殊请求, 触发其与 IPv6 解析器进行合作, 从而发现 IPv6 解析器。为了提高整个过程的挖掘效率, 需要在一次请求中多次实现这一过程。解决这一问题的思路是触发多次查询请求, 而通过 CNAME 记录进行重定向是一种非常有效的方法, 且在前人工作^[5-6]中证明了该方法的有效性。因而, 本文沿用了这一方法构建跨栈域名解析场景, 具体实现方式如下。

(1) 配置子域名

本文的实验需要两个不同的子域作为基础条件, 因此首先对域名记录进行配置。配置域名记录需要注册域名, 从而掌握域名的管理权, 假设注册域名为 a.com。在该域名下, 配置两个不同的子域名, 如 v4.a.com 和 v6.a.com。

(2) 添加子域名的 NS 记录

在确定子域的名称后, 给每个子域各配置一

个独立的 NS 记录, 表明两个子域下的所有域名均由对应的 NS 记录指向的权威服务器负责解析。例如, 将 v4.a.com 的 NS 记录指向 ns4.a.com, 将 v6.a.com 的 NS 记录指向 ns6.a.com。

(3) 添加权威服务器的地址记录

形成跨栈解析的关键在于权威服务器的地址记录的配置。在常见的域名配置中, 权威服务器会配置 A 记录或在双栈情况下配置 A 和 AAAA 记录。而在本文中需要将两台权威服务器分别配置为 IPv4 或 IPv6 单栈模式。例如, 配置一个 A 记录将 ns4.a.com 指向自建权威服务器的 IPv4 地址, 配置一个 AAAA 记录将 ns6.a.com 指向自建权威服务器的 IPv6 地址。

完成上述配置后, 在单栈状态下, IPv4 解析器只能获得 v4.a.com 下域名的资源记录, IPv6 解析器只能获得 v6.a.com 下域名的资源记录。自建权威服务器返回 CNAME 记录, 使得两个子域下的子域名互为对方的 CNAME 记录指向。从而, IPv4 (IPv6) 解析器需要从 IPv6 (IPv4) 权威服务器获取域名记录, 形成跨栈解析。在解析过程中, 通过不断返回 CNAME 记录, 解析器需要解析在两个子域下交替间隔的一系列的域名, 形成 CNAME 记录链条。对应地, 参与到解析过程中的解析器也形成了解析链。结合本文提出的动态域名水印技术, 使 IPv4 与 IPv6 DNS 服务的合作关系暴露在权威服务器的视角, 进而发现 IPv6 DNS 服务地址。

2.2 动态域名水印嵌入

在生成域名资源记录的过程中, 将参与域名解析过程的 DNS 服务关键信息, 即自建权威服务器收到查询的源 IP 地址, 也是本文地址发现的主要目标, 通过特定的编码方式嵌入域名

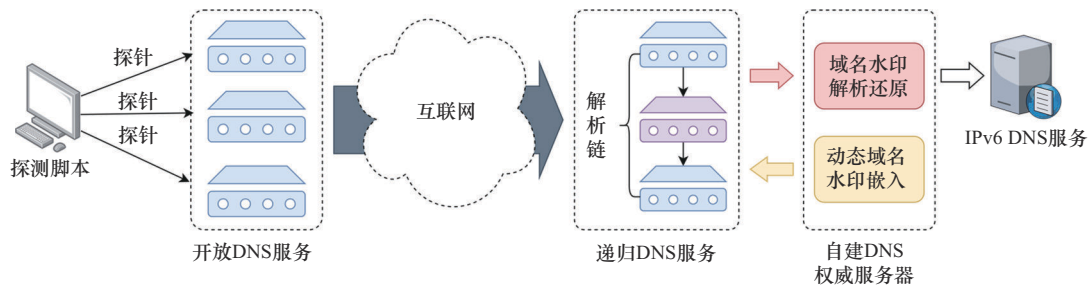


图3 基于动态域名水印的IPv6 DNS服务发现方法的架构

Figure 3 Architectural methodology for measuring IPv6 DNS services based on the dynamic domain name watermark embedding

中,形成域名水印。已嵌入水印的域名可以在客户端或权威服务器端通过对应的解码方式进行水印解析,从而溯源整个域名解析过程。在文献[6]的IPv6 DNS服务发现方法中,解析器将遵循由权威服务器返回CNAME记录形成的解析链进行域名解析;权威服务器负责将解析器的IP地址编码为字符串嵌入CNAME记录中,具体而言是将IPv4的点分十进制(IPv6的冒分十六进制)的分隔符换为“-”,从而满足域名编码要求;当解析器完成整个解析过程后,将通过TXT记录返回完成的水印域名,客户端接收该域名并解析水印,最终发现IPv6 DNS服务。参与CNAME解析链的每一个解析器节点都将对应一个域名水印,因而权威服务器每次返回的CNAME记录中的目标域名的长度将会不断增加。

文献[5-6]方法需要将整个解析链上的解析器通过水印编码嵌入域名中,一旦嵌入则跟随域名解析过程继续向后进行,因此是静态的。虽然这种方法可以较直观地挖掘IPv6 DNS服务的存在并发现IPv4-IPv6 DNS服务关系,但现代DNS复杂的体系结构使得这种方法存在一些弊端。首先,域名字符串长度最大为255个字符。

考虑到IPv4和IPv6地址对应的编码长度,一个域名最多只能嵌入2组域名水印,每组水印包含一个IPv4和一个IPv6地址水印。因此CNAME解析链最长为4,限制了一次请求可以挖掘的IPv6地址数量。其次,文献[11]表明,一些大型的DNS服务商会将同一域名查询交由不同的解析器执行,以提高服务的响应性能,这使得实际的解析过程呈现树状形式。由于最终响应的域名记录中的水印仅能记录链式的解析过程,因此将丢失其他解析器间关系,造成数据损失。

为了解决上述问题,本文方法使用了一种动态域名水印嵌入技术,动态域名水印嵌入的过程如图4所示。该技术与之前工作最大的不同在于将域名水印的嵌入方式由添加水印更改为替换水印,并且将水印解析位置从探测脚本变为了权威服务器侧,其核心思想是只专注于还原解析链相邻两台解析服务器间的关系。当请求到来时,权威服务器首先将请求的源IP编码为水印字符串,并判断域名中是否存在域名水印。若不存在,则将水印字符串嵌入域名中的特定位置并返回CNAME记录;若存在,则将用生成的水印字符串替换原水印,形成新的域名返回CNAME记

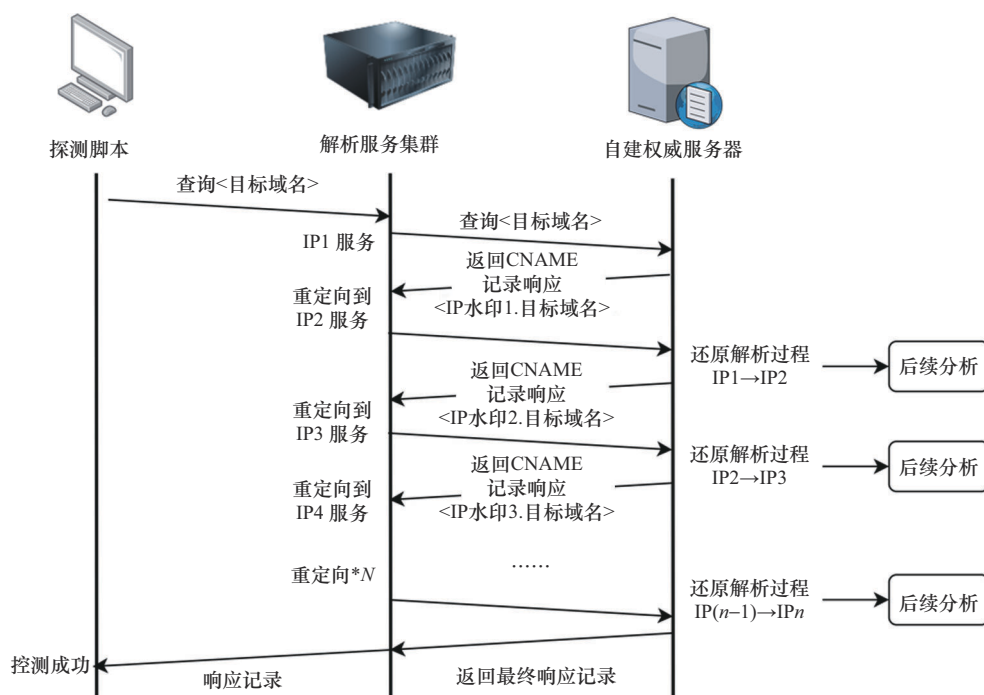


图4 动态域名水印嵌入的过程

Figure 4 The process of dynamically embedding domain name watermark

录。其中, IP 地址的编码方式与之前工作中使用的编码方式相同, 即将 IPv4 的点分十进制 (IPv6 的冒分十六进制) 的分隔符换为 “-”。此外, 由于 CNAME 记录中的域名水印每次都会被新的水印替换, 因此是动态的。一方面, 由于只关注相邻解析器间的关系, 对于之前的解析无须关心, 因此同一时刻域名中只需嵌入一个水印, 在保证水印溯源效果的同时, 缩短了域名长度, 理论上可以无限次地增加 CNAME 解析链的长度。另一方面, 通过收集相邻的解析器关系, 水印嵌入和水印解析是同步进行的, 可以很好地在一次请求中完成对呈树状形式的解析过程的还原, 避免了传统方法中存在的损失, 动态域名水印与静态域名水印的解析过程如图 5 所示。

在向 IPv4 DNS 服务发送 DNS 查询探针之

后, 解析服务集群内部的合作将会被触发并通过 IPv6 DNS 服务向自建权威服务器查询。自建权威服务器只需不断返回 CNAME 记录即可不断触发合作解析并让探测过程持续进行。

本文所提出的 IPv6 DNS 服务发现方法的核心目标在于提高单次探测中可以挖掘的提供 DNS 服务的 IPv6 地址数量。传统基于静态域名水印的方法在探测脚本所在的客户端处执行水印的解析, 因此必须将 IP 编码不断追加到域名中作为静态水印, 直到权威服务器将水印信息作为最终响应返回给探测脚本。这意味着需要等待域名解析过程结束才可以解析水印。并且, 如果需要发掘更多的 IPv6 地址势必增加域名长度, 一些解析器可能认为过长的域名记录存在恶意攻击威胁, 提前终止域名解析。探

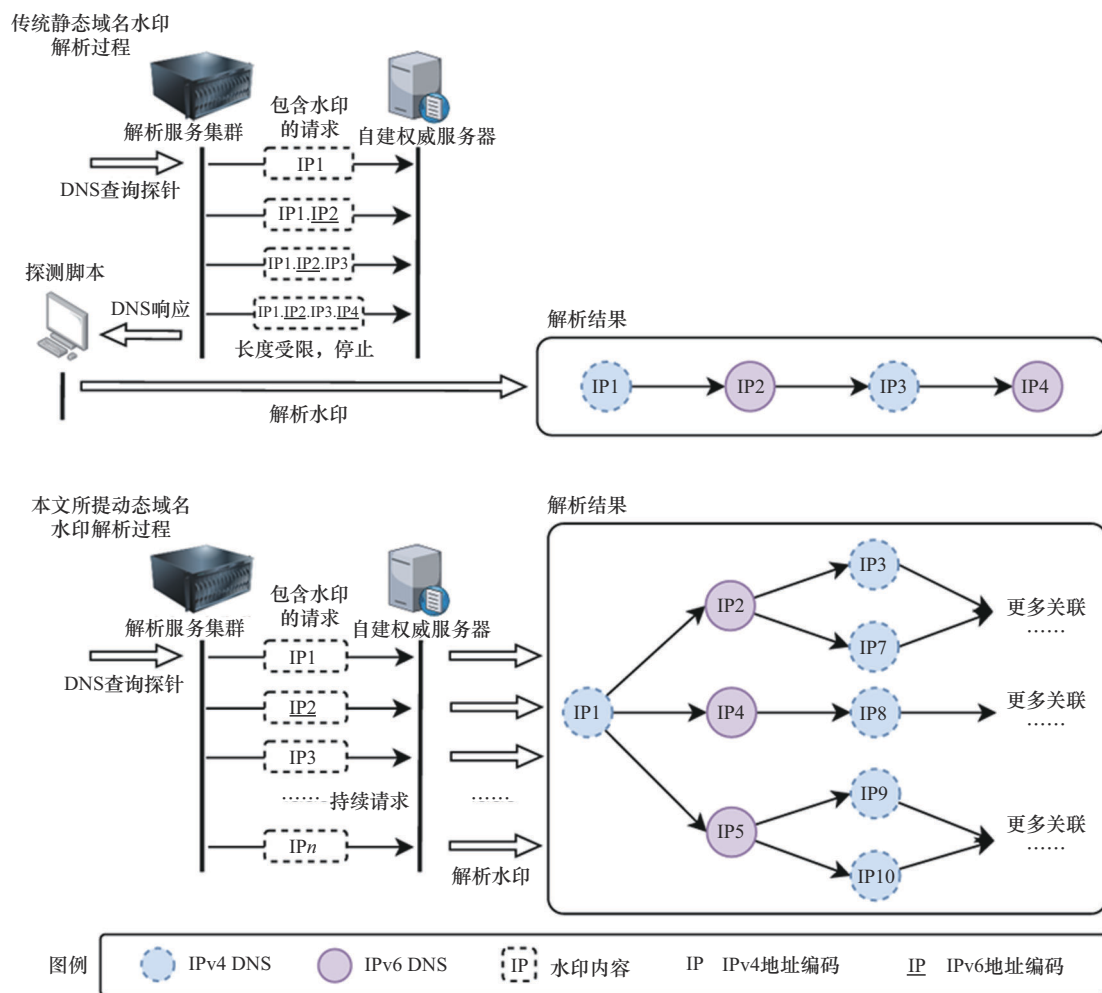


图 5 动态域名水印与静态域名水印的解析过程

Figure 5 Resolution process of dynamic domain watermark and static domain watermark

测脚本将无法收到响应,造成单次探测可还原的IPv6地址数量较有限。而本文方法由于在权威服务侧解析水印,因此无须担心解析器的影响,可以在解析过程中同步解析水印,挖掘服务间关联并发现更多提供DNS服务的IPv6地址。

2.3 IPv6 DNS服务发现

本文研究通过从域名水印中还原IPv4-IPv6 DNS服务间的关联,进而发现IPv6 DNS服务。为了与动态域名水印嵌入过程相对应,本文设计了一种新的域名水印格式作为挖掘IPv6 DNS服务的探测报文,其组成如下。

进度标识.实验对象标识.水印部分.
随机字符.实验域名

进度标识会在每一次的CNAME解析中修改一次,以表示经过的解析链的长度,如c1、c2、c3等。该标识同时可以辅助权威服务器判断是否应当返回TXT记录,以停止解析链,避免解析链无限制增加而对实验目标产生过多网络负担。实验对象标识是探测脚本发包的目的地址的水印编码,用于记录域名解析的IPv4解析器入口。由于DNS系统的复杂性,权威服务器接收到的请求源IP可能与探测脚本发送的实际目标解析器不同。考虑到需要向不同的IPv4解析器发送请求报文,当进行分析时,如果缺乏区分不同入口解析器的标识,就会使数据的

分析变得复杂。水印部分是权威服务器添加或替换的域名水印的位置。随机字符是从36个不同字符(26个英文字母+10个阿拉伯数字)中随机抽取6个形成一个字符串,以增加域名的独特性,避免解析器缓存对实验的干扰。实验域名则决定了该域名是由IPv4还是IPv6权威服务器负责解析。基于该域名结构,探测脚本将会发送包含初始进度标识但不包含水印部分的域名作为探测报文,如

c1.1-2-3-4.random.v4.ourdomain

脚本将查询该域名TXT记录的查询请求发送到IPv4解析器,并由权威服务器负责后续的解析还原与服务发现。IPv4解析器在接收到发送的探针后,将会遵循域名解析过程定位到域名的权威服务器。权威服务器是经过特殊设计的,其将返回两种域名记录的响应:CNAME或TXT记录。探测开始之前,需要为权威服务器配置探测最大长度阈值。权威服务器会记录域名解析跳数,这些域名解析是在一台IPv4开放解析器作为入口的探测中进行的。通过不断返回动态生成的CNAME记录,域名解析过程以链式不断进行下去。当域名解析跳数等于所设定的阈值时,域名解析链到达最后一跳的请求域名将会被用来填充TXT记录并返回给解析器作为最终响应。因此,只有在最后一轮DNS解析进行时才会返回。

权威服务器的处理过程如图6所示,权威服务器将每一个符合格式要求的实验域名的查询请

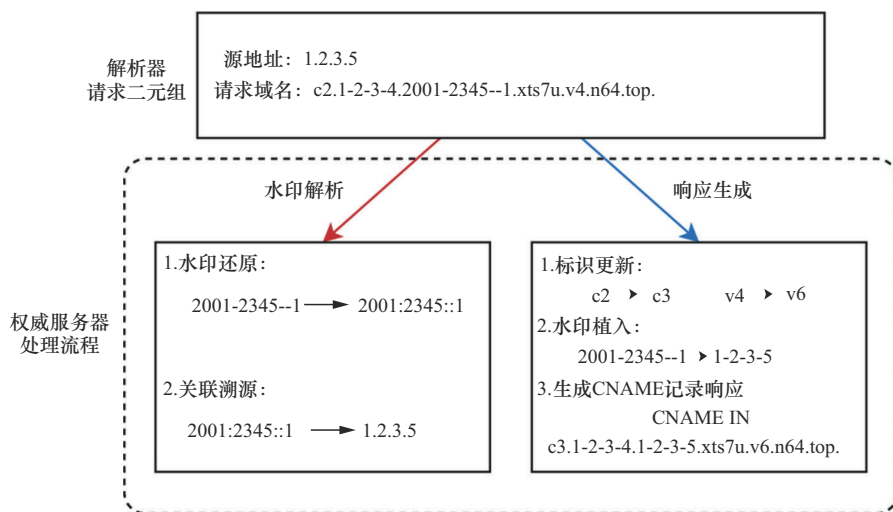


图6 权威服务器的处理过程

Figure 6 The process of the authoritative server handling

求封装为二元组, 其中解析器 IP 即为请求的源 IP。权威服务器将根据接收到的二元组信息进行水印解析以及域名资源记录动态生成。权威服务器首先从域名中提取水印部分, 并还原为正常的 IP 地址, 该地址即为上一跳负责域名解析的解析器的地址, 记为 IP1。将二元组中的解析器 IP 记为 IP2, 即可还原得到一条 IP1→IP2 的解析器地址关联。如前文所述, 权威服务器间隔地返回由单栈 IPv4 或 IPv6 权威服务器负责解析的域名作为 CNAME 记录。因此, 在上述 IP1 和 IP2 中, 必有一个为 IPv6 地址。通过提取这些 IPv6 地址, 即可发现对应的 IPv6 DNS 服务。

3 实验

3.1 实验设置

为了更好地对比本文使用的动态域名水印与静态域名水印的实际探测效果, 需要比较在相同探测条件下权威服务器使用静态域名水印与动态域名水印的实际效果差异。因此, 实验中同时使用了两个域 n64.top 和 n64.live。遵照第 2.2 节所述过程, 实验中对两个域配置了类似的域名记录, 最终使得子域名的解析均由同一台权威服务器负责, 以消除网络环境或硬件配置差异对实验效果造成的影响。

此外, 为了分析自建权威服务器的选择差异对实验的影响, 实验中部署了 3 台硬件配置相同的服务器作为自建权威服务器, 其所处的位置与所属运营商均不相同。为了满足动态域名水印方法的探测条件, 3 台自建权威服务器所负责解析的子域名遵照 2.2 节所述过程配置, 同属于域 n64.top, 但不同权威服务器负责解析的子域名完全不同。因此, 可以通过不同子域名的实验结果分析权威服务器在不同地理位置或不同运营商网络对实验结果的影响。

最后, 实验中使用了自定义探针格式的 zmap 作为探测工具, 以适配本文方法的域名格式。实验中部署了多个探测节点, 这些节点将向全网 IPv4 地址发送目标域名 TXT 记录的查询探针, 以利用开放 IPv4 DNS 解析器挖掘 IPv6 DNS 服务。相关实验环境说明如表 2 所示, 所有服务器配置均为 CPU 两核、内存 4 GB、100 Mbit/s 网络带宽以及 40 GB 存储。

表 2 实验环境说明

Table 2 Description of the experimental environment

部署位置	类型	用途
新加坡	探测脚本	发送域名请求探针
美国纽约	探测脚本	发送域名请求探针
中国杭州	探测脚本	发送域名请求探针
中国香港	自建权威服务器	动态静态域名水印处理与响应生成
日本大阪	自建权威服务器	动态域名水印处理与响应生成
美国纽约	自建权威服务器	动态域名水印处理与响应生成

3.2 实验结果

3.2.1 不同方法测量效果对比

本节将从 IPv6 DNS 服务发现效率、CNAME 链还原长度以及地址关联簇大小维度对比本文提出的基于动态域名水印的探测方法和传统的基于静态域名水印探测方法的测量效果。

本文实验收集了 2023 年 12 月 6 日至 2023 年 12 月 20 日部署在中国香港的自建权威服务器的数据, 其间共完成了两轮次的全球探测。

(1) IPv6 DNS 服务发现效率

评估 IPv6 DNS 服务发现效果, 最直接且重要的方法就是比较不同方法在同一时间段内发现的 IPv6 DNS 服务数量。本文方法和传统方法在同一时间段内发现的 IPv6 DNS 服务数量去重后如图 7 所示, 可以看出, 本文方法相较于传统方法发现的 IPv6 DNS 服务数量提升接近 98%, 表明本文方法在发现 IPv6 DNS 服务方面拥有更好的挖掘效果。

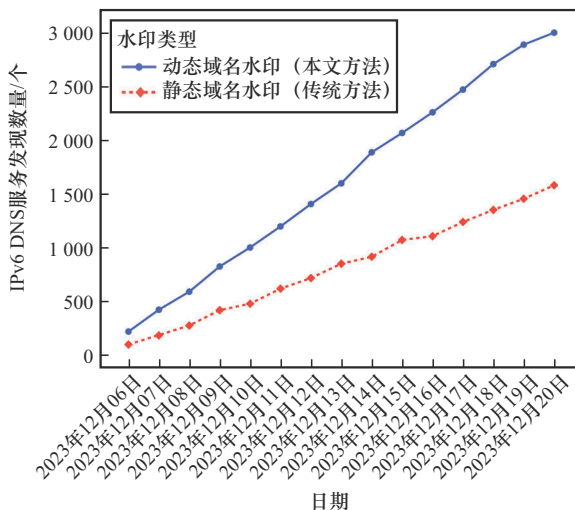


图 7 不同方法的 IPv6 DNS 服务发现数量对比

Figure 7 Statistics on the number of IPv6 DNS services discovered by different methods

这种提升是由以下两个方面原因带来的。

1) 使用动态域名水印可以极大地延长解析链的可用长度, 而无须受域名长度的限制。同时, 由于舍弃了域名中冗余的关联信息, 只关心解析链上一节点与当前节点之间关联的还原, 因此可以降低解析器和权威服务器的网络压力。

2) 本文方法将域名水印的解析过程从客户端转移到了权威服务器侧。由于解析器需要与权威服务器进行通信以完成域名解析过程, 因此从权威服务器的视角可以更好地观察多台出口递归解析器同时处理请求的现象。权威服务器能够同时收集不同的解析链, 形成一棵解析树, 在结果上表现为: 从探测脚本的单个请求开始的解析过程中可以发现更多的IPv6 DNS服务。而在传统方法中, 客户端只能从静态域名水印中还原出一条解析链, 需要依赖后续的重复探测以增加发现概率, 降低了探测的实际收益。

图8从多个方面比较了两种方法的挖掘效率, 与基于静态域名水印的传统方法相比, 本方法利用动态域名水印在每次探测请求中可以还原发现的提供DNS服务的IPv6地址数量提升了接近200%, 探测投入产出比更高。并且, 在每次请求中发现的IPv6地址归属的/64地址块数量上, 也略多于基于静态域名水印的传统方法, 关联挖掘更完整, 可以发现不同地址块中的DNS服务间存在的关联。同时也说明大多数情况下单次探测发现的IPv6地址处于一个地址块中。而在域名的水印部分平均长度上, 本文方法采用替换方式的动态域名水印, 相较于采用追加方式的静态域名水印平均减少了70%, 极大提高了探测过程的传输效率, 避免应答请求的域名过长导致被解析器误判为恶意行为。

(2) CNAME 解析链

IPv6 DNS服务的发现依赖于更全面、完整地还原IPv4-IPv6 DNS服务关联关系。为了更好地解释本文方法带来的探测效率提升的根本原因, 本文分析了依赖于CNAME解析链的服务关联还原结果。由于在域名的格式中添加了进度和实验对象标识, 因此通过权威服务器接收到的域名中的进度标识最大值就可知道对实验对象的单次探测中可触发的CNAME解析链长度, 两种方法收集到的CNAME解析链长度对比如图9所示。

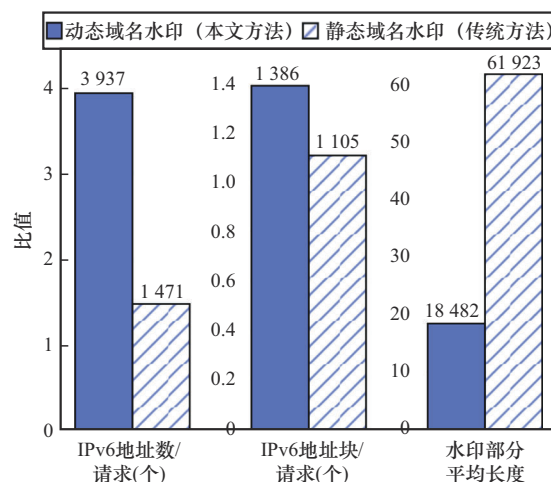


图8 两种方法的挖掘效率对比

Figure 8 Comparison of mining efficiency of two methods

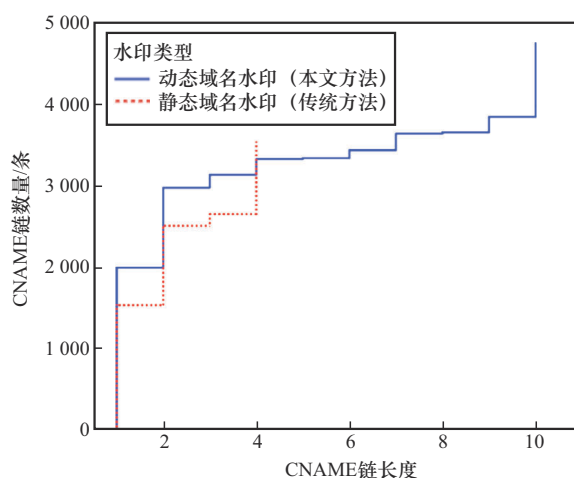


图9 两种方法收集到的CNAME解析链长度对比

Figure 9 Comparison of the distribution of CNAME chain lengths collected by the two methods

实验结果呈现出两个方面的特点。首先, 本文方法成功利用动态域名水印在保证IPv6 DNS服务正常解析溯源的前提下, 形成了更长的解析链。对于一些小型服务商的解析器, 过长的解析链可能会引起请求数量的放大, 增加服务负担。因此, 实验中设置了解析链的最大长度值为10。另外, 部分解析器由于网络环境或者配置问题, 无法进行跨栈解析, 所以解析过程终止, 造成对应的解析链最终长度小于阈值。

其次, 从收集到的CNAME链的整体数量来看, 本文方法相比传统方法提升了约17%。这种提升主要得益于将水印的解析位置从探测脚本代表的客户端侧转移到了权威服务器侧, 从而更好

地适应了多台递归解析器同时请求形成的多解析链现象。

(3) 地址关联簇

本文实验还对还原 IPv4-IPv6 DNS 服务关联形成的地址关联簇的效果进行了分析, 两种方法收集到的地址关联簇大小对比如图 10 所示。地址关联簇指以 IPv4 和 IPv6 DNS 地址作为节点、地址关联作为边形成的节点关联图。能够收集形成更大的地址关联簇意味着对实际网络环境中 IPv4 和 IPv6 DNS 服务关联关系的感知更精确, 同时意味着 IPv6 DNS 服务的发现能力更强。

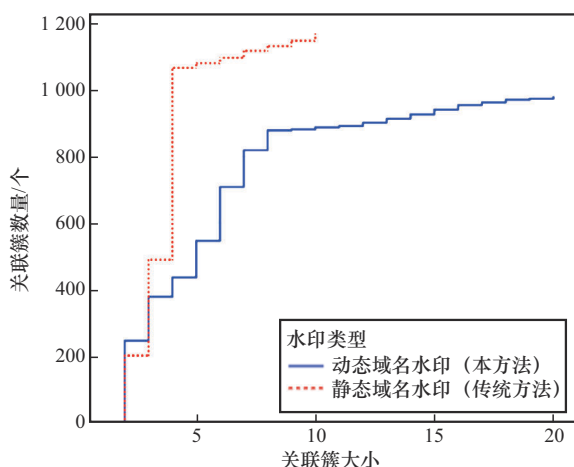


图 10 两种方法收集到的地址关联簇大小对比

Figure 10 Comparison of the distribution of address correlation cluster sizes collected by the two methods

从图 10 不难看出, 相比传统方法, 本文方法发现的地址关联簇平均规模更大。这主要是因为本文方法在探测脚本的一次请求中可以挖掘更多的服务间的跳转关系, 从而更加完整地还原域名解析过程, 而无须重复扫描。因此, 将得到的关联关系进行整合后, 得到的地址关联簇规模更大。其次, 相比传统方法, 本文方法挖掘到的关联簇的数量更少。这种现象的原因是传统方法得到的不同关联簇之间存在未被发现的地址关联, 而基于动态域名水印的本文方法由于探测挖掘服务的能力更强, 发现了未被前者发现的两个关联簇中的联系, 最终合并为一个规模更大的关联簇, 造成整体数量减少。

此外, 在安全性方面, 相比传统方法, 本文提出的动态域名水印方法能降低探测过程中域名的平均长度, 避免应答与请求的域名过长导致被解析器误判为恶意行为, 具有更高的隐蔽性。同

时, 解析链的长度限制也能避免增加服务器的解析负担。

3.2.2 不同权威服务器测量效果对比

为了分析自建权威服务器的选择对发现 IPv6 DNS 服务数量的影响, 本节对比了 3 台不同地理位置的自建权威服务器在相同两个探测周期内的实验数据。3 台自建权威服务器的地址位置和归属运营商均不相同, 因此可以比较判断权威服务器在不同国家及网络环境中对收集 IPv6 DNS 服务数量的影响。不同自建权威服务器收集到的 IPv6 地址数量统计如图 11 所示。

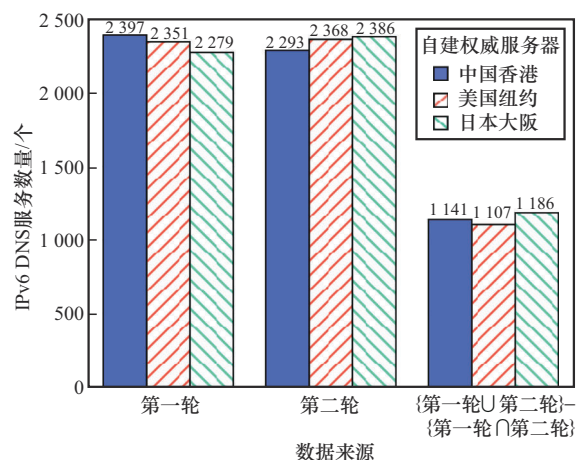


图 11 不同自建权威服务器收集到的 IPv6 地址数量统计

Figure 11 Statistics on the number of IPv6 addresses acquired by diverse self-constructed authoritative servers

从统计结果可以发现, 不同的自建权威服务器在同一轮次的探测中发现的 IPv6 地址数量存在 2% 到 5% 的差异; 相同权威服务器在两轮次中收集到的 IPv6 地址数量接近。根据调研, 服务商的负载均衡策略可能造成了地址集合的差异。例如, Xu 等^[12]的工作成果表明, 一些服务商虽然拥有大量可用的解析器集群 (即解析器池), 但在一次域名解析过程中使得一台或数台解析器参与其中。同时, 由于 DNS 的复杂性, 同一域名的查询请求并不一定由相同的解析器负责处理。因此, 单轮次探测并不能完全挖掘解析器池中的所有解析器。对于本文提出的 IPv6 DNS 服务发现方法而言, 根据服务商的负载均衡策略, 向相同的 IPv4 地址发送请求, 实际负责域名解析的 IPv6 DNS 递归出口 IP 可能不同。本文在多轮次实验中也观测到在长期持续探测下

IPv6 地址数据集合与已发现的 IPv6 地址重合度逐渐增大的现象，佐证了这一结论。综上所述，实验结果表明，发现的 IPv6 DNS 服务数量主要取决于持续探测的时间跨度，与权威服务器所处网络环境或地理位置无明显关联。

3.2.3 IPv6 DNS 服务发现效果

聚焦到基于动态域名水印的 IPv6 DNS 服务发现方法自身的数据，本文分析了 2023 年 12 月 06 日至 2024 年 2 月 5 日通过向全网开放 IPv4 DNS 服务发送域名查询探针得到的所有 IPv6 DNS 递归解析器的数据。实验数据收集过程共经过 11 轮次的探测，得到 22 519 个不同的承载 DNS 服务的 IPv6 地址。由于利用了 IPv4 解析器关联发现 IPv6 解析器，本文方法发现的所有 IPv6 解析器地址都曾实际向自建权威服务器发送过 DNS 查询。这种网络行为是解析器组件所特有的。因此，本文方法发现的所有 IPv6 地址均部署有 DNS 服务。

为了更好地了解实验发现的 IPv6 DNS 服务的具体情况，对于实验中收集到的承载 DNS 服务的 IPv6 地址，利用 IP 信息库进行了 IP 信息的查询匹配，并对结果进行了统计分析。这些结果将展现 IPv6 DNS 服务在网络空间与地理空间中的分布特点。在网络空间中，承载着 DNS 服务的不同 IPv6 地址由运营商控制的自治域进行管理。因此，本文首先统计了 IPv6 DNS 服务在不同网络自治域中的分布情况，可视化统计结果如图 12 所示。

图 12 (a) 的结果表明，目前 IPv6 DNS 服务的分布并不均衡。约有 40% 的自治域仅存在 1 个 IPv6 DNS 服务。而部分自治域拥有的 IPv6 DNS 服务数量超过 2 000 个。本文分析了数量排名前 10 的自治域拥有的 IPv6 DNS 服务数量，如图 12 (b) 所示，可以发现 IPv6 DNS 服务在网络空间中的分布具有明显的头部效应。ASN15169 和 ASN3462 拥有的 IPv6 DNS 服务数量超过 2 000 个，与紧随其后的自治域 ASN7713 相比是后者数量的将近 3 倍。这种现象说明不同自治域的 IPv6 DNS 服务发展较悬殊。并且，观察前 10 个自治域的管理组织，可以发现大多数为大型的全球网络服务商，这表明 IPv6 DNS 服务的部署和服务运营商本身的服务规模和运营体量有一定关系。

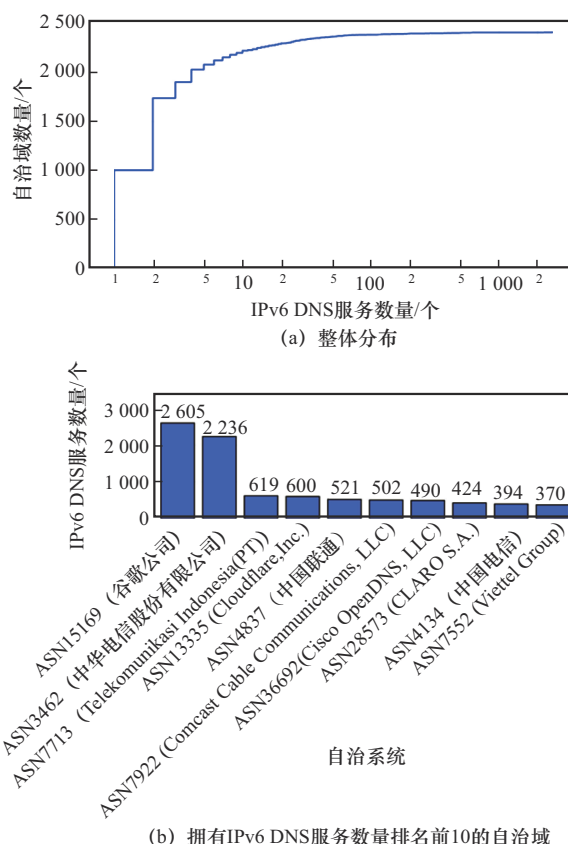


图 12 不同自治域 IPv6 DNS 服务分布统计
Figure 12 Statistics distribution of IPv6 DNS services in different AS

由于 DNS 服务对互联网发展具有重要的作用，因此大部分递归解析器是由运营商负责搭建和维护的。分析不同运营商拥有的 IPv6 解析器数量可以了解当前 IPv6 DNS 服务生态的发展特点，评估不同运营商提供 IPv6 域名解析服务的能力，并挖掘 IPv6 DNS 服务发展的主要推动力量。不同运营商的 IPv6 DNS 服务分布统计如图 13 所示。

首先，发现的 IPv6 DNS 服务涉及的运营商总数与自治域总数差距不大，表明大多数运营商基本仅在一个管辖的自治域内部署 IPv6 DNS 服务。因此，与自治域的数据特点类似，接近 50% 的运营商仅拥有一个 IPv6 DNS 服务。部分运营商则拥有超过 2 000 个不同的 IPv6 DNS 服务。图 13 (b) 展示了拥有 IPv6 DNS 服务数量排名前 10 的运营商。这些运营商均为知名电信服务商或 DNS 服务商，说明当前 IPv6 DNS 服务存在明显的头部聚集效应，规模较大的运营商往往拥有较强的 DNS 服务能力。

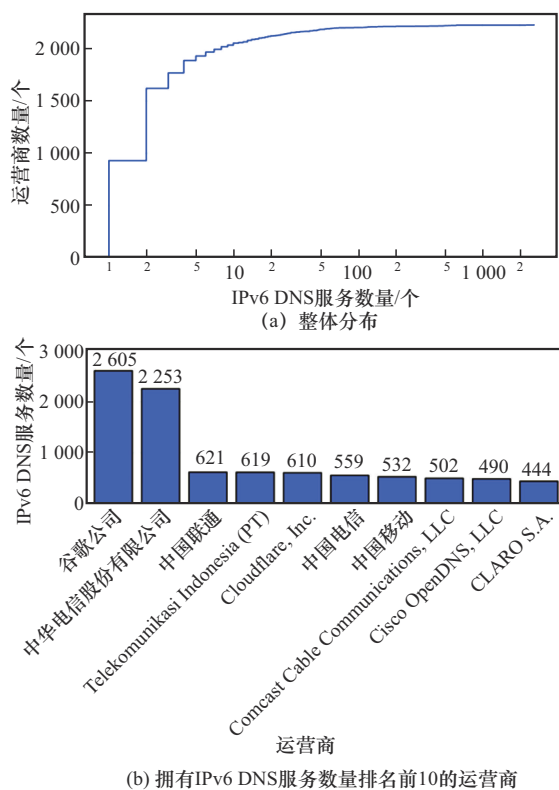


图 13 不同运营商的 IPv6 DNS 服务分布统计

Figure 13 Statistics distribution of IPv6 DNS services in different ISPs

IPv6 DNS 解析器不仅是运营商维持自身域名系统运行的重要组件，还可以辅助国家更好地进行网络管理和信息安全维护，如防止网络攻击、过滤不合适的内容等。IPv6 DNS 服务的地理分布能在一定程度上反映不同国家的 IPv6 发展状态。基于 IPv6 DNS 服务对于国家网络发展的重要作用，本文统计了 IPv6 DNS 服务在不同国家的分布情况，从统计数据来看，目前已发现的 IPv6 解析器主要集中分布在中国和美国，在欧洲和南美洲地区的一些国家分布也较广泛。然而，在相当多非洲地区的国家并没有发现 IPv6 DNS 服务。这说明目前 IPv6 DNS 的部署发展进度与各地区经济发展状况有一定关联。同时，为了结合实验结果分析不同国家的域名基础设施的发展情况，本文利用公开数据集统计了 IPv6 开放解析器的分布情况。根据公开数据集 IPv6 Hitlist 中开放 UDP/53 端口的活跃 IPv6 地址的地理位置信息，分析了 IPv6 开放解析器在全球的分布情况。根据公开数据的统计可视化结果，IPv6 开放解析器主要分布在中国和美国。将公开

数据与本文实验所发现的 IPv6 DNS 服务数据进行比较，发现两个数据集结果大体一致。值得注意的是，不同数据集中关于中美两国的 IPv6 DNS 服务部署数量存在一定差异。这一现象的主要原因在于两个数据集探测方法的不同。对于 IPv6 DNS 服务发现而言，本文结果是通过自建权威服务器收集得到的，主要是递归解析器的地址；而 IPv6 Hitlist 数据集通过主动探测方法得到，主要是开放解析器的地址。探测方法的不同造成两个数据集中关于中美 IPv6 DNS 服务的数量出现不一致的现象。

除了探测得到的 IPv6 DNS 服务本身的分布情况，本文还针对收集到的服务关联进行了统计分析。从地址间关联的角度出发，本文发现一些承载 DNS 服务的 IPv6 地址与大量不同的 IPv4 地址存在关联关系。存在这一现象的 IPv6 地址往往隶属于大型的 DNS 服务商，图 14 展示了谷歌某一 IPv6 DNS 服务节点与其他 DNS 服务的关联关系。

该 IPv6 DNS 服务与大量的 IPv4 DNS 服务存在着关联关系，显著增加了该节点的重要性。一旦该节点被恶意地注入错误的缓存记录，就会同时影响所有存在关联的 IPv4 DNS 服务。同时，一些攻击者也可能利用这些 IPv4 DNS 服务构造放大攻击，给该 IPv6 DNS 服务节点造成极大的网络和计算压力，触发拒绝服务攻击。

4 结束语

针对现有工作中使用静态域名水印挖掘 IPv6 DNS 服务受域名长度限制，造成发现效率低的问题，本文提出利用动态域名水印技术进行 DNS 解析过程还原，以提高 IPv6 DNS 服务的挖掘效率。本文方法首先在跨栈解析场景下构建 CNAME 解析链，在一次请求中形成多次跨栈解析，迫使 IPv4 解析器委托 IPv6 解析器向权威服务器查询；然后，运用动态域名水印技术，不断替换域名水印内容以记录解析链中下一跳的解析器 IP 信息；最后，在权威服务器中解析请求，发现 IPv6 解析器。与前人基于静态域名水印的方法相比，本文方法可绕过域名长度对水印记录方面的限制，形成更长的解析链，触发更多的解析器参与到域名解析中，增加了探测收益；同时，

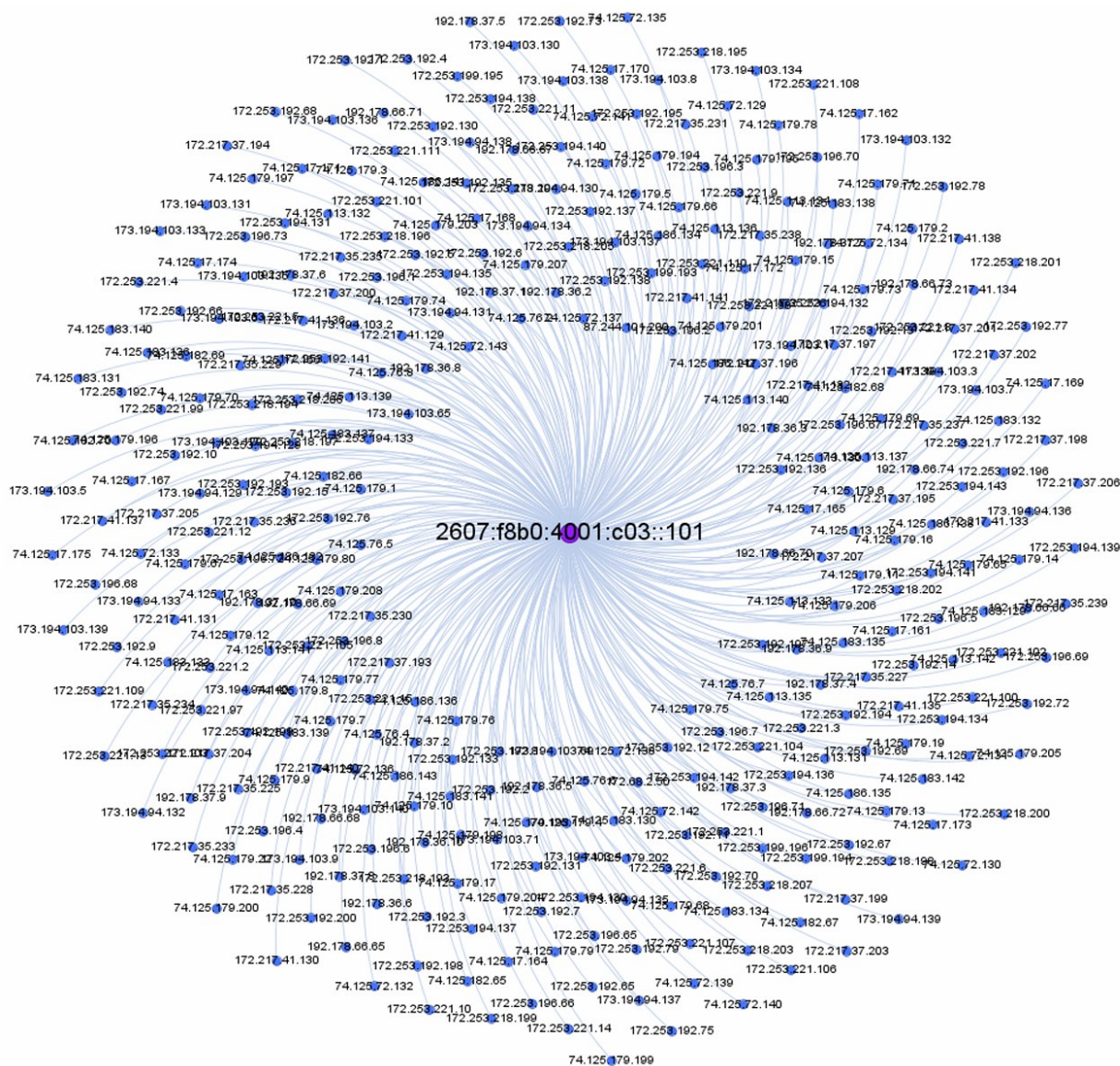


图14 谷歌某一IPv6 DNS服务节点与其他DNS服务的关联关系

Figure 14 The association between a specific Google IPv6 DNS service node and other DNS services

由于水印的解析过程位于权威服务器侧，与传统方法在DNS客户端解析水印相比，可以更加完整地溯源域名解析过程，尤其是在多台出口解析器同时解析域名的情况下。通过与基于静态域名水印的对比实验、对实验周期内收集到的IPv6 DNS服务的数据分析，证明了本文方法可以实现对IPv6 DNS服务的有效挖掘。最后，本文方法利用探测得到的DNS服务之间的关联，挖掘得到了大量解析依赖和集中化现象信息，为分析DNS服务在IPv6环境下的安全态势提供了数据基础。

参考文献：

- [1] CUI T, XIONG G, GOU G, et al. 6VECLM: language modeling in vector space for IPv6 target generation[C]//Proceedings of the Machine Learning and Knowledge Discovery in Databases: Applied Data Science Track: European Conference. Berlin: Springer, 2021: 192-207.
- [2] CUI T, GOU G, XIONG G, et al. 6GAN: IPv6 multi-pattern target generation via generative adversarial nets with reinforcement learning [C]//Proceedings of the IEEE INFOCOM 2021-IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2021: 1-10.

- [3] LIU Z, XIONG Y, LIU X, et al. 6Tree: efficient dynamic discovery of active addresses in the IPv6 address space[J]. Computer Networks, 2019, 155: 31-46.
- [4] YANG T, HOU B, CAI Z, et al. 6Graph: a graph-theoretic approach to address pattern mining for internet-wide IPv6 scanning[J]. Computer Networks, 2022, 203: 108666.
- [5] BERGER A, WEAVER N, BEVERLY R, et al. Internet nameserver IPv4 and IPv6 address relationships[C]//IMC '13: Proceedings of the 2013 Internet Measurement Conference. New York: Association for Computing Machinery, 2013: 91-104.
- [6] DURAND A, IHREN J. DNS IPv6 transport operational guidelines [R]. 2004.
- [7] FIEBIG T, BORGOLTE K, HAO S, et al. Something from nothing (there): collecting global IPv6 datasets from DNS [C]//Proceedings of the Passive and Active Measurement: 18th International Conference. Berlin: Springer, 2017: 30-43.
- [8] BAGNULO M, SULLIVAN A, MATTHEWS P, et al. DNS64: DNS extensions for network address translation from IPv6 clients to IPv4 servers[R]. 2011.
- [9] AL-DALKY R, SCHOMP K. Characterization of collaborative resolution in recursive dns resolvers[C]//Proceedings of the Passive and Active Measurement: 19th International Conference, Berlin : Springer, 2018: 146-157.
- [10] MO D, ZHU Y, JIE Z, et al. Unveiling flawed cache structures in dns infrastructure via record watermarking[C]//GLOBECOM 2023-2023 IEEE Global Communications Conference. Piscataway: IEEE Press, 2023: 6024-6029.
- [11] MOURA G C M, CASTRO S, HARDAKER W, et al. Clouding up the internet: how centralized is DNS traffic becoming?[C]//Proceedings of the ACM Internet Measurement Conference. 2020: 42-49.
- [12] XU C, ZHANG Y, SHI F, et al. Measuring the centrality of DNS infrastructure in the wild[J]. Applied Sciences, 2023, 13(9): 5739.

[作者简介]



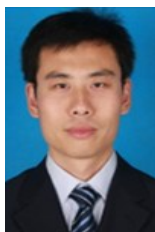
韩丁康 (1999-), 男, 河南焦作人, 中国科学院信息工程研究所硕士生, 主要研究方向为网络空间测绘。



朱宇佳 (1984-), 女, 江苏无锡人, 博士, 中国科学院信息工程研究所副研究员, 主要研究方向为网络空间测绘、域名安全、网络大数据分析。



赵蕾 (2001-), 女, 山东泰安人, 中国科学院信息工程研究所硕士生, 主要研究方向为网络空间测绘。



焦亮 (1986-), 男, 山东滨州人, 中国科学院信息工程研究所博士生, 主要研究方向为网络空间测绘。



刘庆云 (1980-), 男, 河北人, 博士, 中国科学院信息工程研究所正研级高级工程师、博士生导师, 主要研究方向为网络空间测绘、网络空间安全。