DNS 根服务体系的发展研究

延志伟[1,2]，耿光刚[1,2]，李洪涛[1,2]，李晓东[1,2]
（1. 中国互联网络信息中心，北京 100190；
2. 互联网域名管理技术国家工程实验室，北京 100190）

摘　要：以 DNS 根服务体系发展历史为切入点，阐述了根服务器及根区文件的管理模式，针对 DNS 根服务器管理模式面临的效率、可扩展性以及稳定性等方面的缺陷，从政策层面综述了国际社群对扩展 DNS 根服务器的相关讨论和结论，并基于若干开放性原则进一步从技术层面详细论述和分析了扩展 DNS 根服务体系的相关解决方案。

关键词：域名服务；根系统；任播；泛在 DNS 根服务
中图分类号：TP391
文献标识码：A
doi: 10.11959/j.issn.2096-109x.2017.00150

Study on the development of the DNS root system

YAN Zhi-wei[1,2], GENG Guang-gang[1,2], LI Hong-tao[1,2], LI Xiao-dong[1,2]
(1. China Internet Network Information Center, Beijing 100190, China;
2. National Engineering Laboratory for Internet Domain Name Management, Beijing 100190, China)

Abstract: The development history of the DNS root system was described and the management of the DNS root service was explained in detail. Based on the shortcomings on efficiency, scalability and stability of the current DNS root server management model, the extension schemes from both policy and technology points of views were summarized and analyzed.

Key words: domain name, root system, anycast, ubiquitous DNS root service

1　引言

在互联网蓬勃发展的今天，互联网用户迅猛增加，各种上层应用层出不穷。域名服务系统（DNS，domain name system）作为解析互联网资源名字及互联网资源地址的基础服务，其重要性愈发突出。而作为 DNS 解析入口的根服务体系，其安全稳定是整个域名解析业务正常高效运作的先决条件。

DNS 根服务器用于响应用户对根区文件（root zone file）的查询请求，根区文件维护着顶级域名（TLD，top level domain）的位置信息，全球共有 13 台根服务器。1997 年 8 月，1 台根服务器被从美国转移到日本，13 台根服务器的格局基本形成（除了位于日本的 1 台外，9 台位于美国，欧洲的 2 台分别位于英国和瑞典）。

由于 DNS 所使用的传输协议——用户数据报协议（UDP，user datagram protocol），对数据分组具有 512 B 的长度限制，要让所有的 DNS 根服务器信息被包含在同一个 UDP 数据分组中，根服务器数量只能被限制为 13（准确地说，13 台根服务器所需的 DNS 响应数据分组大小为 436 B），且每个服务器要使用字母表中的单个字母

收稿日期：2016-12-12；**修回日期：**2017-02-11。**通信作者：**耿光刚，gengguanggang@cnnic.cn
基金项目：国家自然科学基金资助项目（No.61375039, No.61303242）
Foundation Item: The National Natural Science Foundation of China (No.61375039, No.61303242)

（A~M）标识。13 台服务器由 12 个独立机构运维（其中 VeriSign 运维 2 台根服务器），这些机构起初都是以自愿者身份被选出。此外，出于 DNS 根服务多样性考虑，这 12 个机构均按照自身规划和模式管理对应的根服务器。当前 13 台根服务器的基本信息如表 1 所示。

美国东部时间 2002 年 10 月 21 日下午，这 13 台服务器遭受了有史以来最为严重的也是规模最为庞大的一次分布式拒绝服务（DDoS，distributed denial of service）攻击。超过常规数量 30~40 倍的数据量猛烈地向这些服务器袭来，从而导致其中的 9 台不能正常运行。事后，DNS 根服务体系开始采用任播（anycast）技术进行 DNS 根服务的复制，到 2004 年，DNS 根服务器镜像节点已多达 80 台，它们分布在 34 个不同的国家和地区。2007 年 2 月 6 日，DNS 根服务器再次遭受大规模 DDoS 攻击，攻击持续了近 8 h，攻击源几乎遍布全球，该攻击事件发生后，DNS 根服务器镜像已增加到 130 台，分布在 53 个国家和地区。近几年的一次根服务器大规模扩展发生在 2012 年，著名黑客组织 Anonymous 在 2012 年 2 月宣称要采用放大和反射攻击摧毁 DNS 根服务体系，这一事件促使 DNS 根服务体系在短短几个月内几乎在全世界各国都部署镜像，其总量超过了 300 台。截至 2016 年 12 月 5 日，13 台根服务器通过任播技术在全球部署服务节点已达 641 个。

中国在 2003 年引进了第一个根服务器的镜像——F 根镜像，是由 ISC 和中国电信共同建立的。2005 年，I 根的管理机构 Autonomica 在中国互联网络信息中心（CNNIC）设立了中国第二个根镜像。2006 年，中国网通与美国 VeriSign 公司合作，正式开通 J 根的中国镜像服务器。2011 年，CNNIC 在北京新增一个 F 根镜像。此外，CNNIC 于 2012 年又部署了第一个 L 根镜像节点。2014 年，世纪互联、北龙中网和天地互连 3 家公司分别与互联网名称与数字地址分配机构（ICANN，Internet corporation for assigned names and numbers）开展合作，在中国增设 3 台 L 根域名服务器镜像节点。阿里云与 VeriSign 合作在杭州建设了一个 J 根镜像。这 4 个根服务器的 9 个镜像节点成为我国境内 DNS 查询请求主要的根服务节点。

因此，从历史发展看，DNS 根服务体系（本文所用“根服务体系”包括根区文件管理体系与根服务器管理体系）随着互联网的不断繁荣，越

表 1 **根服务器主要情况**

根服务器	运维机构	IP 地址	AS 号码	镜像数量
A	VeriSign, Inc.	IPv4: 198.41.0.4 IPv6:2001:503:BA3E::2:30	19836	4
B	Information Sciences Institute	IPv4: 192.228.79.201 IPv6: 2001:478:65::53	4	0
C	Cogent Communications	IPv4: 192.33.4.12	2149	7
D	University of Maryland	IPv4: 128.8.10.90 IPv6: 2001:500:2D::D	27	108
E	NASA Ames Research Center	IPv4: 192.203.230.10	297	70
F*	ISC	IPv4: 192.5.5.241 IPv6: 2001:500:2f::f	3557	57
G	U.S. DOD NIC	IPv4: 192.112.36.4	5927	5
H	U.S. Army Research Lab	IPv4: 128.63.2.53 IPv6:2001:500:1::803f:235	13	1
I*	Autonomica	IPv4: 192.36.148.17 IPv6:2001:7fe::53	29216	49
J*	VeriSign, Inc.	IPv4: 192.58.128.30 IPv6: 2001:503:C27::2:30	26415	117
K	RIPE NCC	IPv4: 193.0.14.129 IPv6: 2001:7fd::1	25152	46
L*	ICANN	IPv4: 199.7.83.42 IPv6: 2001:500:3::42	20144	157
M	WIDE Project	IPv4: 202.12.27.33 IPv6: 2001:dc3::35	7500	7

注：“*”表示在中国境内具有该服务器镜像节点。

来越成为各国政府机构、学术研究机构和产业界普遍关注的热点，围绕 DNS 根服务体系的扩展与架构演进的讨论多年来一直在延续，如今随着互联网全球化管理模式的讨论再次成为关注焦点。

本文对 DNS 根服务体系的发展历史进行了梳理，并阐述了 DNS 根服务体系的管理模式，对扩展 DNS 根服务体系的方案进行了分析，且对其未来的演进方向进行了探索。

2　DNS 根服务体系管理模式

DNS 通过层次化的形式管理域名数据，从而以分段的方式将人们可以记住的域名转换为计算机使用的数字以寻找其对应的目的地。DNS 管理体系中最为核心的角色便是互联网数字分配机构（IANA，Internet assigned numbers authority），其职能由 ICANN 承担。

ICANN 是一个非盈利组织，总部设在美国，此前一直按照与美国商务部（US DoC/NTIA）签订的谅解备忘录（MOU）来行使与互联网码号资源管理相关的职能。除了与美国商务部签订了谅解备忘录之外，ICANN 还按照其与商务部签订的一项单独合同行使 IANA 的职能（这些职能具体包括根区管理、协调技术协议参数的分配以及互联网码号资源的分配）。图 1 为从 ICANN 角度理解的 IANA 职能管理权移交前 DNS 根区管理模式；而在根区管理过程中，ICANN 首先需要向 NTIA 提交来自国际顶级域名（gTLD）和国家代码顶级域名（ccTLD）注册管理机构对根区文件的修改申请，NTIA 批准修改后，再授权运维隐藏根的 VeriSign 公司进行对应的操作。修改后的文件最后会经由隐藏根向 13 个根服务器进行扩散，进而扩散到全球的根镜像节点。

鉴于 ICANN 和美国政府在 DNS 根区文件管理中所扮演的重要角色，业界一直认为美国政府和 ICANN 对 DNS 根服务体系这一互联网重要基础设施存在绝对的控制权。准确而言，这种认识是有偏颇的：为了保证根区文件的一致性，当前采用以 ICANN 协调、NTIA 批准、VeriSign 操作的集中模式（ICANN 为 IANA root zone management function operator，VeriSign 为 root zone maintainer，NTIA 为 root zone administrator）。虽然 ICANN 也不断优化这一管理模式，如在 2013 年对外发布了根区管理的审计报告，但这模式在一定程度上有悖互联网多利益相关方（multi-stakeholder）的原则：一方面，ICANN 同时承担了政策制定与操作实施的角色，对 TLD 可用性认定和具体操作实施没有功能的分离；另一方面，由于美国政府涉入被视为国家资源的 ccTLD 的修改操作（尽管

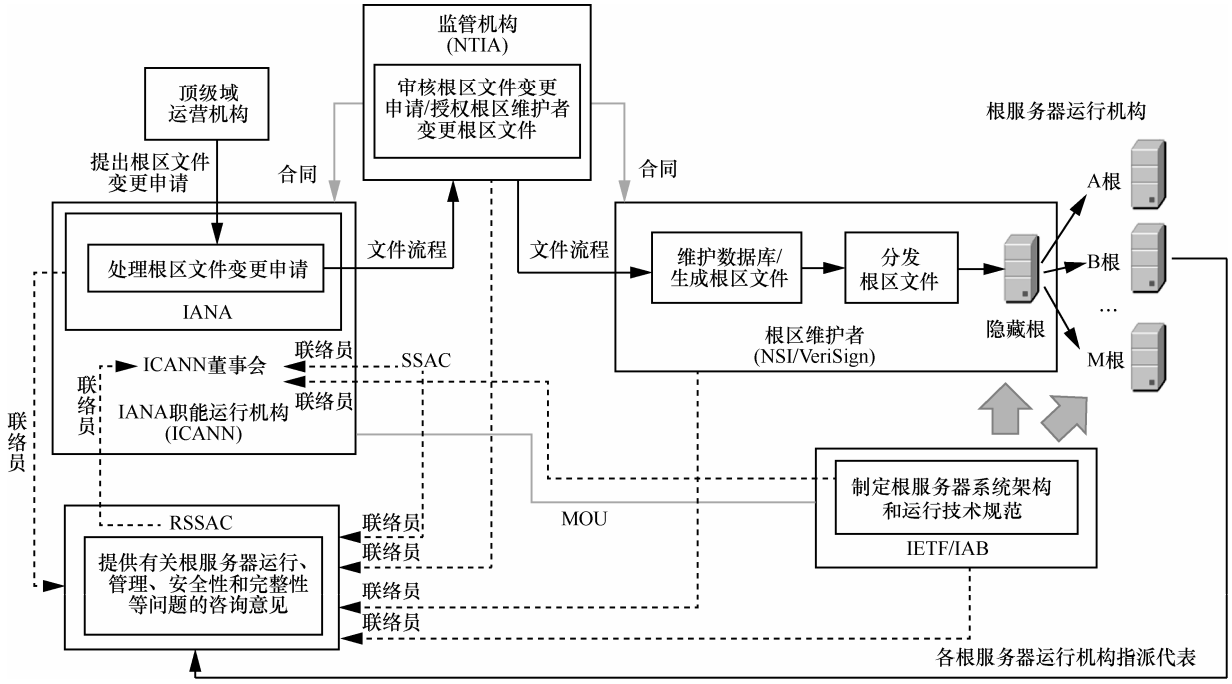


图 1　域名体系管理模式

NTIA 认为其角色仅是监管操作流程是否合规进行），因此业界普遍认为此模式并不足够公平开放。这一问题也是在 IANA Transition 进程中重要的议题，如社群广泛建议应进一步明确根区管理的安全保证和外部审计机制、提升 ICANN 政策制定透明度、推动现有咨询讨论的公开透明化，以强化全球互联网社群的参与和监督。

从根服务器的角度看，每个 DNS 根服务器运行机构才拥有对该服务器的绝对管理权，但并不存在对所有根服务器具有集中管理权利的实体（在 2012 年 ICANN 与 NTIA 签订的更新版合同中也明确 ICANN 对 DNS 根服务体系的管理限于根区（DNS root zone management）层面，而非以前较为含糊的 DNS 监管（domain name system supervision，1997 年的合同）或 DNS 根的监管（administrative functions associated with root management，2000 年的合同）等说法）。这种分布式管理模式旨在保证 DNS 根服务的部署多样性（diversity）和运行稳定性（stability）。实际上，DNS 根服务器架构的多样性保证也体现在 DNS 根服务器所使用的软件上，为了避免单点失效和运行风险，这些根服务器尽量采用多样化的 DNS 软件并使用不同的版本。

在 IANA 职能管理权移交后，美国政府将不再承担相关审核和监督职责，全球根管理合同基本格局将由 ICANN 主导的 2 份合同确定：一是与移交后 IANA（PTI）签订 IANA 域名职能合同，授权 PTI 履行 IANA 职能运行工作；二是与 VeriSign 签订根区维护者服务协议（RZMA），授权 VeriSign 继续负责根区文件修改、生成、维护全球根区数据库系统以及根区文件的分发。这 2 份合同已分别于 2016 年 10 月 1 日和 2016 年 10 月 20 日生效。

由此可见，根区文件管理的优化更多是国际政策层面的问题，本文重点从技术层面研究探讨 DNS 根服务器运行管理模式存在的问题及可行的演进方向和扩展机制。

3 DNS 根服务器扩展分析

对根服务器数量是否应该随着域名解析系统的发展突破当前限制，以更加开放、可扩展的方式增设，国际互联网协会（ISOC，Internet society）早年便有过对此问题的考虑，但其认为 DNS 根服务器的当前架构应以稳定性为主，不宜轻易做出改动。无独有偶，这个问题在国际电信联盟（ITU，International Telecommunication Union）也讨论过，结论也很明确。ITU 认为增加根服务器并非受阻于技术问题，主要是对其分配和管理很难抉择。然而，随着互联网的发展，各个国家和地区都有在本国本地区增设根服务器的期望，以此来强化对互联网核心基础设施管理的参与度并优化 DNS 根解析的本地服务性能。

2002 年，日本研究人员对根服务器的数量和分布进行过研究，认为当时的根服务器分布严重不均，希望能对欧洲和美国的根服务器进行重新部署[1]，不过其背景是尚未采用 anycast 技术部署镜像节点[2,3]。随着互联网的不断发展，DNS 根服务器随着 DNS 的演进承载更多的功能，同时基于 anycast 技术在规模上不断扩张，ICANN 也意识到未来不断扩展根服务器的需求和可能性，因此在 2009 年 2 月的董事会决议中要求根服务器咨询委员会（RSSAC，root server system advisory committee）、安全和稳定咨询委员会（SSAC，security and stability advisory committee）和 ICANN 工作人员深入研究引入 IPv6 地址记录、国际化顶级域名（IDN，internationalized domain name）、其他新的通用顶级域名，以及为支持 DNS 安全扩展协议（DNSSEC，domain name system security extensions）在根区增加新的资源记录对 DNS 根服务器的稳定性影响[4]。作为对该决议的答复，RSSAC、SSAC 和 ICANN 工作人员成立了根扩展指导性工作组（RSSG，root scaling steering group）来为此次集中性的研究制定研究范畴和预计研究成果，并将其公布于 2009 年 5 月的参考条目（ToR）中[5]，该指导性工作组还成立了一个专家组（RSST，root scaling study team）着手该项研究[6]。作为主要成果，RSST 于 2009 年 8 月和 2009 年 10 月相继发表了 2 篇研究报告，名为《扩展根：关于扩展根区规模及增大根区波动对 DNS 根系统影响的报告》[7]和《根扩展研究：DNS 根扩展模型的描述》[8]。前者认为根区文件承载的内容越来越多，势必对根服务器的稳定性

造成影响，特别是对网络状态较差的区域，在根区文件更新频繁的时候可能会存在一定困难，进一步估算认为，当前的数据同步架构所能承受的每年新增根区文件数据的条目在 $O(100)$量级，如若上升到 $O(1\ 000)$的量级，势必需要对当前的根服务器架构进行适应性调整；后者给出一个量化的根区数据管理模型，可用于仿真和评估根区扩展研究相关的问题。

除了参与 RSST 相关议题外，RSSAC 还就如何高效、安全、稳定地管理 DNS 根服务器提出若干建议，重点声明根服务器运行机构仅是根区文件发布者（publisher）而非修订者（editor）[9]；此外，应切实强化根服务器运行机构的可审计性并制定运维管理的相关准则和服务水平规范[10]；并表明根服务器作为 DNS 解析的入口，应及时更新相关功能支持（如 TCP、IPv6 等），保证根区文件的准确性及最大限度增强根服务器部署的泛在性[11]。

这些来自 ICANN 的研究和建议不仅从另一个角度对扩展当前的 DNS 根服务器体系提出实际需求，也在一定程度上为未来 DNS 根服务器体系的扩展奠定了技术和政策基础。结合相关研究报告，本文从 DNS 根服务器运行的性能角度分析其部署架构，可发现影响其安全稳定及服务性能的具体因素主要体现在 2 个方面：根区文件的同步延时以及 anycast 节点造成的 BGP 路由收敛开销。本节首先对这 2 个方面进行分析，并进一步从实际运行状态角度阐述 DNS 根服务器架构的缺陷。

3.1　根区文件同步时延

当前的根区文件更新操作由 VeriSign 负责，其频率为一天 2 次，具体流程如图 2 所示[8]。当新的根区文件产生后，分发主体（DM，distribution master），即上文提及的隐藏根，向所有的其他根服务器（RS，root server）发送 DNS 通告消息（notify），每个 RS 相应地回复确认消息（acknowledgement）。如果 DM 没有在规定时间内接收到确认消息，将会重新发送通告消息，尝试与 RS 建立联系。RS 成功发送确认消息之后，随即向 DM 发送起始授权记录（SOA，start of authority）请求，以此来验证自己当前的根区文件版本与 DM 所维护的根区文件版本之间是否存在差异。DM 以当前根区文件序列号进行响应。如果 DM 响应的版本号大于 RS 当前维护的根区文件的版本号，RS 则启动区传输（XFR，zone transfer）以请求更新根区文件。

当采用 anycast 技术进行根服务器节点镜像复制时，根区文件也可能采用相同的机制在镜像节点进行同步。然而，由于不断扩大的镜像节点部署规模，这一机制也具有不断增大的时延和开销：第一阶段是 notify 交互、SOA 查询以及 XFR 启动过程；第二阶段是区文件数据传输过程。当根服务器与其镜像节点之间距离较远时，文件同步时间会线性增加，此外，随着根区文件的增大，数据同步时间也会相应增大，这一因素会受到很多 DNS 扩展技术的影响，如 DNSSEC 会将传统区文件扩大 7~10 倍[12]，而 IPv6 资源记录的引入会给每个域名带来额外的 128 bit[13]。另一方面，新的 gTLD 的不断扩张也使根区文件不断增大[14]。

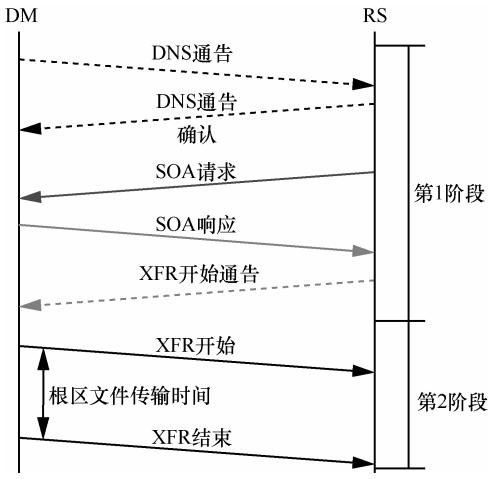


图 2　DM 和 RS 之间的交互流程

正如 ICANN 所分析的，当前有些根服务器运维机构已经发现某些远端镜像节点在根区文件同步中暴露出效率问题，进一步认为，随着根区的不断扩大会给根区文件的同步带来更大挑战[7]。

3.2　BGP 路由收敛开销

研究表明，很多互联网故障归咎于路由聚合的延迟以及路由表的振荡，骨干网路由尤其明显，其平均聚合时间可达几分钟，这也是边界网关协议（BGP，border gateway protocol）路由长久以来广受诟病的原因之一[15]。基于距离矢量算法，BGP 需要每个路由器维护到达可能目的地的距离以及下一跳的向量。当网络连接状态发生变化时，路由器

需要重新计算到目的地的距离以更新路由表[16]。由于 anycast 完全依赖于 BGP 选择最优节点，BGP 收敛的问题自然也影响到基于 anycast 的 DNS 根服务器服务稳定性[17, 18]。如果网络状态不稳定或 BGP 路由属性误配置，都有可能造成 DNS 根区解析服务的性能下降[19]，此外，BGP 路由不断进行动态调整和变化，如果 DNS 承载于传输控制协议（TCP，transmission control protocol）之上，还可能造成同一会话的不同数据报文被路由到不同镜像节点的情况，从而导致 DNS 会话中断。

3.3　解析性能探测与分析

为了直观展示我国境内访问 DNS 根服务器的性能，从而发现其可能的缺陷，本文在全国 32 个省市部署了 61 个监测节点，以探测当前在我国境内部署的 DNS 根服务器镜像节点的运行情况[20]。图 3 为从两大互联网服务提供商 ISP（Internet service provider）（中国电信和中国联通）探测的 13 个 DNS 根服务器的平均解析时延。

由此可见，在国内部署镜像节点的服务器具有较小的时延，但不同服务器节点的差异较大，这一结果从侧面说明：当某个服务器失效或不可达时，该区域的 DNS 根区解析效果将受到较为显著的影响。

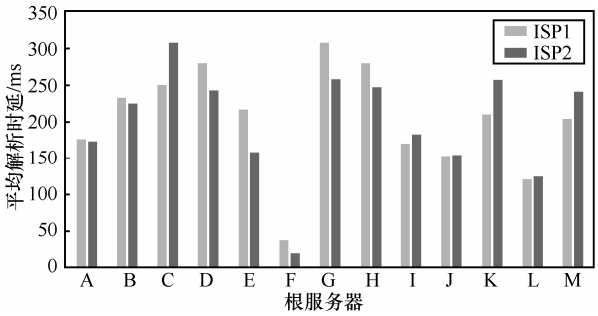


图 3　根服务器平均解析时延

在我国部署镜像的根服务器中，F 节点性能最优，图 4 为从国内主要省份访问 F 节点的性能。

因为监测节点并未能在所有省份完全覆盖 2 个 ISP，所以图 4 混合了 2 个 ISP 的探测结果（如在河北只有 ISP2 网络的监测节点）。上述结果表明，虽然不同位置具有较大差异，但 F 节点的解析时延整体较低（大多低于 50 ms），这是因为大部分访问 F 节点的请求都命中部署在国内的 F 镜像节点。

如图 5 所示，“pek2a”和“pek2b”是国内 F 镜像节点所使用的 2 个 IP 地址的标识，10 次测

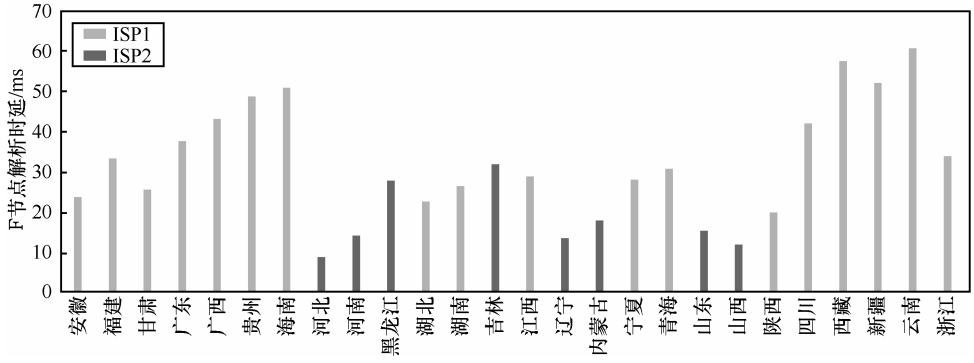


图 4　F 节点解析时延

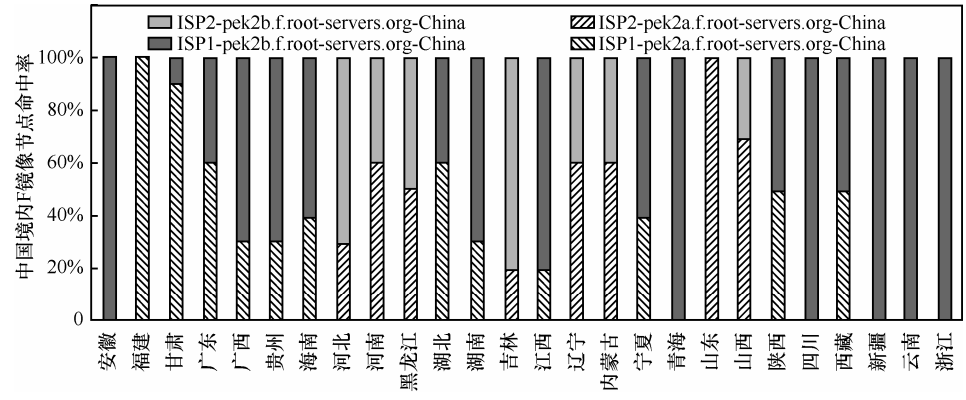


图 5　中国境内 F 镜像节点命中情况

量都命中这 2 个 IP 地址。

相比于 F 节点，J 节点的解析时延明显增大，如图 6 所示，大多数访问的时间都超过 150 ms。相应地，图 7 给出了访问 J 节点命中国内镜像的情况。

图 7 中“elbei1”标识 J 在国内镜像节点所使用的 IP 地址，“v6sfol”为 J 在美国 San Francisco 的镜像节点所使用的 IP 地址。当请求消息命中国内镜像节点时，时延明显较低，如在安徽和山东的监测结果都低于 50 ms。但是其他省份的监测请求都命中了美国的镜像节点，时延明显增大。

上述探测结果受到该地区部署递归服务器的数量、性能及与国内镜像节点之间距离和 ISP 在该地区链路状态等诸多因素的影响，但也可以直观地发现，在我国网络覆盖范围较广的情况下，集中式部署（几乎均在北京）的 9 个镜像节点显然不能为超过 6 亿互联网用户提供高效稳定的 DNS 根解析服务[21]。从这个角度看，DNS 根服务器的数量扩展和部署优化确实存在很大空间。

4　根服务器扩展原则

虽然当前存在很多关于根服务器演进方案的讨论，但要保证 DNS 根服务器架构能健康演进，首先需要从原则上进行研究和分析。回顾整个互联网发展进程，由于其遵循去中心化（decentralization）、本地化（locality）、可扩展性（scalability）等多种根本性原则，才得以枝繁叶茂，长久繁荣[22]。那么 DNS，特别是其根服务器架构是否在发展过程中良好地遵循这些原则？

1) 去中心化

去中心化保证了互联网控制的民主，增强了错误容忍[23]。在 DNS 体系中，递归服务器层面完全遵循这一原则，因此，递归服务也是 DNS 服务体系发展最为迅速的一环，并侧面推动了整个 DNS 的发展。而根以下的权威服务可以被任何 DNS 区所用（甚至是私有区），这表明顶级及以下权威服务层面在一定程度上也符合去中心化的原则。但 DNS 根服务器自诞生之初就由 12 个不同的运营机构管理，虽然根区文件在理论上应该

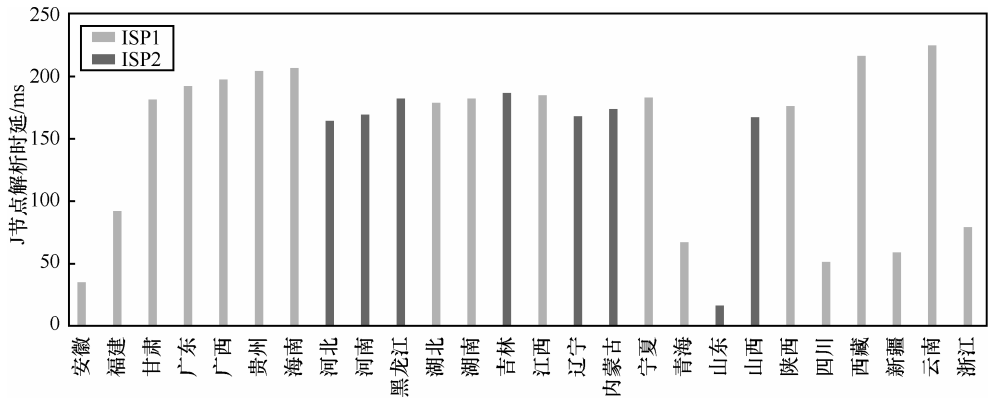


图 6　J 节点解析时延

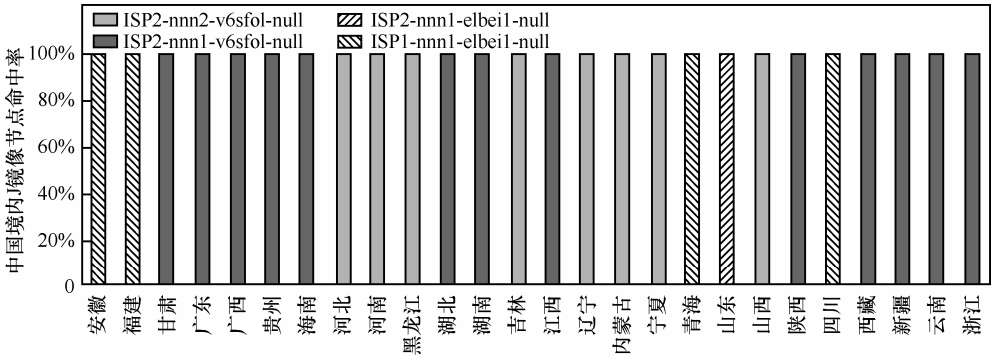


图 7　中国境内 J 镜像节点命中情况

保证唯一性，但对于根服务器的运行管理却没有中心化的必要。

2) 本地化

本地化可以使互联网的失效被限制在本地范围[24,25]，从而增强互联网的整体生存性和健壮性。从DNS服务体系看，递归服务器可以根据本地业务实际需求进行本地化部署，很好地遵循了这一原则。类似地，从顶级域及其以下，数据安全性受到DNSSEC的保障，并不需要服务器集中性部署和管理。但这一原则并未在DNS根服务器上得到体现，虽然DNS根服务器采用了anycast技术，但是镜像节点的管理受到上级节点的严格制约，这种自上而下的模式很显然是对DNS根服务器本地化和多样性需求的最大束缚。

3) 可扩展性

可扩展性保证了互联网可以任意发展和扩张[26]。正如前所述，由于遵循去中心化和本地化原则，DNS递归服务器和顶级及以下权威服务器具有良好的可扩展性。但DNS根服务器的可扩展性却受制于BGP路由体系，正如前所述，这也造成实际运行及未来发展的若干问题。

由此可见，设计去中心化的、本地化的以及不严格依赖于其他协议和服务体系的DNS根服务器扩展机制才是保证DNS根服务器架构顺应互联网发展理念的可行演进思路。

5 DNS根服务器扩展方案

对于未来扩展DNS根服务体系的可行方向考虑，当前可分为2种思路：在当前13个根服务器基础上新增根服务器并弥补13个根服务器在地理位置分布不均等方面存在的缺陷[27~29]；设计能满足长远需求的开放DNS根服务体系架构。在此基础上，本文结合第4节的演进原则提出一种泛在DNS根服务体系及其关键技术。

5.1 服务器数量扩展

DNS是一个分布式系统，所有的查询在缓存没有命中的情况下都是从根区开始的，因此递归服务器必须配置根服务器的地址，作为查询的入口，这个配置文件称之为根区提示文件（hint file），该文件包含所有根服务器的名字和对应的IP地址。递归服务器管理员可以从指定位置下载，同时递归服务器每一次启动后，都会根据配置的根区提示文件，向其中一个根服务器查询根服务器授权记录以及Glue记录（即服务器IP地址）来更新可能更改的根服务器信息，这个过程被称为Priming，探测的过程是使用UDP发送查询请求。所以为了完成一次探测，应答分组应获得所有的根授权记录和对应的Glue记录，并以此作为以后查询根区信息的依据。

在没有任何IPv6记录之前，根配置了13台权威服务器，每个服务器有一条Glue记录，整个应答分组大小为436 B，而IPv4网络上最保守（基于路径MTU（PMTU，path maximum transmission unit）安全值）的UDP报文大小限制在512 B内，这也是当初设计13台根服务器时主要考虑的因素。

随着IPv6协议的引入，根服务器开始配置使用IPv6的地址，从而造成Priming应答数据分组长度突破512 B。例如在7台根服务器上配置IPv6地址，Priming应答消息将会增大到63 B，这就超过了IPv4中安全UDP报文限度值，一次探测应答的报文中，将包含所有的IPv4的地址，而只能包含2条IPv6的地址，返回哪2个根服务器的IPv6地址，不同的服务器可以有不同的实现，有的根服务器实现是不区分IPv4和IPv6地址的，即返回部分IPv4地址和部分IPv6地址，这就导致被返回IPv6地址的根服务器，潜在地接受更多的基于IPv6的DNS查询。

由于这个缺陷，DNS递归服务器软件在BIND9之后开始启用EDNS0（extension mechanisms for DNS version 0）[30]扩展协议，通过在递归服务器和权威服务器之间协商和探测能支持的UDP分组大小，来增大UDP分组的最大限制以容纳整个应答。CNNIC的探测结果表明，当前递归服务器对EDNS0的支持率已经高达98%[20]。因此，随着IPv6的普及，如果所有的根服务器都已配置了IPv6地址[31]（当前已有11台支持IPv6），13台根服务器信息的报文总长度为811 B（包含一个11 B的OPT记录）[32]。基于RIPE的测试和统计，目前使用的绝大部分的递归服务器都能支持811 B以上的UDP报文，而且目前使用中的大部分网络中间设备都允许大于512 B的UDP报文通过。进一步地，2010年7月，根区数据将被

DNSSEC 签名，使每个 DNS 查询的应答中都会包含新的签名记录，超过 512 B 已经是非常普遍的事情。为此，RFC4035[33]指出，支持 DNSSEC 的服务器必须支持 EDNS0。这就表明随着互联网的发展，已经不能再将 DNS 报文小于 512 B 作为无法增加新的根服务器的阻碍。同时，RFC5966[34]还要求 DNS 服务器支持 TCP 查询。一次 TCP 应答的最大长度是 65 535 B，在这种情况下，再增加新的根服务器对报文长度的影响就变得更小。而 CNNIC 探测结果表明，当前递归服务器对于 TCP 查询的支持率也已经达到 74%[20]。

由上述分析可见，当前增设新的根服务器从技术上是完全可行的，这些新增的服务器为后续更加分布式、平衡地部署 DNS 根服务器创造了可能。

5.2　服务模式优化

为了优化 DNS 根服务器架构，当前 DNS 根服务器的运行模式也随着国际社区全面推进的 IANA Transition 被提上议程。ICANN 的 RSSAC 成立了 Caucus[35]，作为其重点工作，Caucus 就 DNS 根服务器安全、稳定以及未来演进的相关问题进行深入研究并向 ICANN 提供技术性咨询。与此同时，ICANN 还就当前 DNS 根服务器运行的模式、根服务器运行机构的审计、准入和推出机制征集了广泛的社群意见[36]。

同时，互联网工程任务组（IETF，Internet engineering task force）也从技术角度开展了相关问题的讨论。DNSOP（domain name system operations）工作组[37]提出了 2 种解决方案，分别从递归服务器一侧和权威服务器一侧实现 DNS 根服务器的扩展。前者[38]主张通过在递归服务器的本地环回接口（loopback）上维护根区文件以实现 DNS 根服务的本地化；而后者[39,40]则通过开放当前 13 台根服务器中的某个（或多个）服务器的地址或通过增设第 14 台开放根服务器，实现根服务器的开放 anycast，以优化根服务节点的可扩展能力。从功能和性能上对比，前者弱化了 DNS 根服务器的重要性，并能实现 DNS 根区解析性能最大程度的优化，且避免了当前大量无效请求影响根服务器整体运行性能的问题。但该方案首先无法保证所有递归服务器运营机构有能力提供和维护这一服务，其次对传统递归服务器运行逻辑改造较大。后者适合灵活部署和推广，但是仍然依赖于 anycast 技术[41]，所以大量的 DNS 根服务器节点可能由于配置不当对 BGP 路由体系造成较大影响[42,43]。虽然这 2 个方案的共同点是均依赖于 DNSSEC 技术保证根区文件同步的安全性及弱化根区文件的来源权威性，但根区文件同步的效率是这 2 个方案共同存在的核心难题。

基于在递归服务层面实现本地化根服务这一方案在实际部署中存在的问题，CNNIC 又提出了共享缓存的解决方案，该方案通过在自治范围内或多个自治范围间共享根区文件缓存服务器，实现根区文件解析的本地化，经过广泛调研，这一机制也是当前很多递归服务提供机构实际采用的运作模式[44]，但由于 DNS 整个生态体系存在扭曲，这种工作模式并未被正式讨论和规范。

此外，还有很多文献提出了采用对等网络（P2P，peer-to-peer）实现分布式根服务器管理架构[45]以及区域性对等根服务器架构（alternative DNS root）[46]，但由于此类研究仅存在于理论层面或有损互联网平等互联原则[47]，考虑到 DNS 根服务对整个互联网安全和稳定的特殊作用，这些方案并不足够成熟以进行实际和大规模广泛部署。

5.3　泛在 DNS 根服务体系

结合 DNS 根服务体系演进原则以及当前解决方案的方向，本文提出一种泛在 DNS 根服务体系，如图 8 所示。

基于 DNSSEC，根区文件的完整性和正确性有了保障，因此，DNS 根服务可以由 DNS 服务体系中的任何逻辑功能来承担（本文将这种可能部署在任何逻辑功能上的 DNS 根服务器称为泛在 DNS 根服务器），但传统的 13 台 DNS 根服务器及其镜像节点、新增的 DNS 开放根服务器及其本地镜像节点、递归服务器的 loopback 接口、甚至顶级及以下的权威服务器。这种架构不仅能满足 DNS 根区解析的性能最优，而且最大程度地保证了 DNS 根区服务的可靠性。

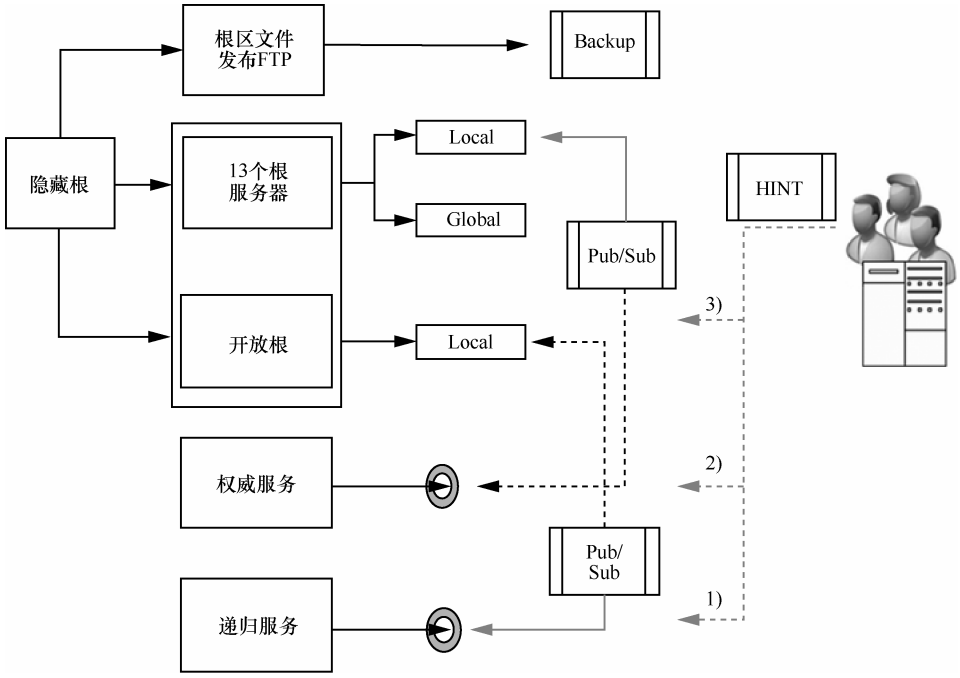


图 8 泛在 DNS 根服务体系

显然，在泛在 DNS 根服务体系中，DNS 服务的部署已经不是问题，但 DNS 服务的泛在化带来的 2 个重大挑战是 DNS 根服务的宣告和根区文件的同步：前者保证了泛在 DNS 根服务能够被用户配置和使用；而后者保证了大量泛在 DNS 服务器能够高效地得到最新的根区文件。

1) 基于 HINT RR 的服务宣告

为了规范化泛在 DNS 服务器配置，本文提出一种新的 DNS 资源记录，称为 HINT，其格式如下。

Zone Lifetime IN HINT Server-name

Zone 标识这个泛在 DNS 根服务器的作用范围，如 CN 标识在中国范围，baidu.com 标识在百度的内部网络。

Lifetime 标识这个资源记录的有效生存期。

IN 标识是一条互联网类型（Internet class）的资源记录。

HINT 标识这条资源记录用于记录该区域内的泛在 DNS 根服务器。

Server-name 为提供该泛在 DNS 根服务器的服务器名称。

递归服务器如果需要配置泛在 DNS 根服务器，就查询对应区的 HINT 资源记录，并将其相应数据加入 db.root 文件中，作为该递归服务器查询根服务器的启动文件。那么递归服务器将可以采用如下 2 种具体策略使用 13 台服务器之外的其他 DNS 根服务器。

① db.root.global.with.local：泛在 DNS 根服务器与传统 A-M 根混用，这是本文建议采用的默认方案，当泛在根服务器不可用时，可以迅速切换到传统的 DNS 根服务器。

② db.root.only.local：单独维护和启用泛在 DNS 根服务器。

2)主被动混合的根区文件同步

如图 8 所示，除了传统根服务器的文件同步方式外，泛在 DNS 根服务体系中的根区文件同步可采用 2 种模式：泛在 DNS 根服务器主动经由根区文件发布点（如 FTP）进行文件下载和更新、采用基于 Pub/Sub 的被动接收方式。其中，前者适用于服务范围较小的泛在 DNS 根服务器，从而可以减轻 DNS 根区文件发布点的压力；后者适用于服务范围较大的泛在 DNS 根服务器，从而可以保证根区文件能在更新后最短时间内得到更新。而 FTP 站点、任何传统的 DNS 根服务器以及开放根都可以作为 DNS 根区文件发布站点，为了避免大量泛在 DNS 根服务器部署造成的根区文件发布瓶颈效应，建议采用层次方式进行数据同步。

6　结束语

作为支撑互联网正常运作的核心基础服务，DNS 随着互联网在普及广度和应用深度的双重驱动下凸显着越发重要的作用。随着 IANA 职能转移（IANA transition）的完成，DNS 根服务体系如何顺势演进也成为整个互联网社区关注的焦点。

本文首先对 DNS 根服务体系的演进历史和当前的运行和管理模式进行了介绍，然后分析了当前 DNS 根服务器运行和管理架构面临的效率、可扩展性以及稳定性方面的问题，并针对中国的网络环境，实际探测了主要省份的根服务性能，从侧面表明需要对当前 DNS 根服务器进行优化扩展的实际需求。

此外，本文从互联网演进的角度出发，提出了 DNS 根服务器未来演进应遵循的若干原则，总结了当前业界讨论的若干方案且进行了分析比较，在此基础上提出了一种泛在 DNS 根服务体系并对关键问题给出针对性解决方案。

参考文献：

[1] LEE T, HUFFAKER B, FOMENKOV M, et al. On the problem of optimization of DNS root servers' placement[C]//Passive and Active Network Measurement Workshop (PAM). 2003.

[2] HARDIE T. Distributing authoritative name servers via shared unicast addresses[S]. IETF RFC3258, 2002.

[3] PARTRIDGE C, MENDEZ T, MILLIKEN W. Host anycasting service[S]. IETF RFC1546, 1993.

[4] ICANN Board of Directors. Draft minutes of the special board meeting[R]. 2009.

[5] ICANN Root Scaling Steering Group (RSSG). Root scaling study terms of reference[R]. 2009.

[6] ICANN. Report of the security and stability advisory committee on root scaling[R]. 2010.

[7] ICANN. Scaling the root-report on the impact on the DNS root system of increasing the size and volatility of the root zone[R]. 2009.

[8] ICANN. Description of the DNS root scaling model, TNO information and communication technology[R]. 2009.

[9] ICANN Root Server System Advisory Committee (RSSAC). Draft proposal, based on initial community feedback, of the principles and mechanisms and the process to develop a proposal to transition NTIA's stewardship of the IANA functions[R]. 2014.

[10] ICANN. Service expectations of root servers[R]. 2013.

[11] ICANN. RSSAC recommendation on measurements of the root server system[R]. 2014.

[12] [EB/OL]http://www.cisco.com/web/about/ac123/ac147/archived_issues/ipj_ 7-2/dnssec.html.

[13] ICANN. Accommodating IP version 6 address resource records for the root of the domain name system[R]. 2007.

[14] CAIDA. Analysis of the DNS root and gTLD nameserver system: status and progress report[R]. 2008.

[15] PERLMAN R. Interconnections[R]. 1999.

[16] GARCIA-LUNA-ACEVES J J. Loop-free routing using diffusing computations[C]//IEEE/ACM Trans. on Networking. 1993: 130-141.

[17] LABOVITZ C, AHUJA A. Delayed Internet routing convergence[C]//IEEE/ACM Trans. on Networking. 2001: 293-306.

[18] LABOVITZ C. The impact of internet policy and topology on delayed routing convergence[C]// IEEE Infocom, 2001.

[19] SARAT S, PAPPAS V, TERZIS A. On the use of anycast in DNS[C]//IEEE ICCCN. 2006.

[20] 中国互联网络信息中心.中国域名服务安全状况与态势分析报告[R]. 2014.
China Internet Network Information Center. Chinese domain name service security situation and trend analysis report[R] .2014.

[21] 中国互联网络信息中心. 第 35 次中国互联网络发展状况统计报告[R]. 2015.
China Internet Network Information Center. The 35th China Internet network development state statistic report[R]. 2015.

[22] CARPENTER B. Architectural principles of the Internet[S]. IETF RFC1958, 1996.

[23] LIMONCELLI T A, HOGAN C J, CHALUP S R. The practice of system and network administration[R]. 2007.

[24] DENNING P J. The locality principle[J].ACM Communication, 2005, 48(7):19-24.

[25] Future Internet Architecture (FIArch) Group. Future Internet design principles[R]. 2012.

[26] CLARK D, CHAPIN L, CERF V, et al. Towards the Future Internet Architecture[S] .IETF RFC1287, 1991.

[27] ABLEY J. Hierarchical anycast for global service distribution[R]. ISC technical note 2003-1, 2003.

[28] SAVAGE S. The end-to-end effects of internet path selection[C]//ACM Sigcomm. 1999.

[29] SPRING N, MAHAJAN R, ANDERSON T. Quantifying the causes of path inflation[C]//ACM Sigcomm. 2003.

[30] VIXIE P. Extension mechanisms for DNS(EDNS0)[S]. IETF RFC2671. 1999.

[31] VIXIE P, KATO A, ABLEY J. DNS response size issues[R]. 2014.

[32] ARENDS R. Protocol modifications for the DNS security extensions[S]. IETF RFC4035. 2005.

[33] BELLIS R. DNS transport over TCP-implementation requirements[S]. IETF RFC5966. 2010.

[34] DEERING S, HINDEN R. Internet protocol, version 6 (IPv6) Specification[S]. RFC2460. 1998.

[35] [EB/OL]https://www.icann.org/resources/pages/rssac-caucus-2014-05-06-en.

[36] ICANN. Overview and history of the IANA functions[N]. 2014.

[37] [EB/OL]http://tools.ietf.org/wg/dnsop.

[38] KUMARI W, HOFFMAN P. Decreasing access time to root servers by running one on loopback[R]. 2015.

[39] LEE X D, VIXIE P, YAN Z W. How to scale the DNS root system?[R]. 2014.

[40] OHTA M. Distributing root name servers via shared unicast addresses[R]. 1999.

[41] SATO S, MATSUURA T, MORISHITA Y. BGP anycast node requirements for authoritative name servers[R]. 2007.

[42] BUSH R, KARRENBERG D, KOSTERS M, et al. Root name server operational requirements[S]. IETF RFC2870, 2000.

[43] Identifying and characterizing anycast in the domain name system[R]. USC/ISI Technical Report. 2011.

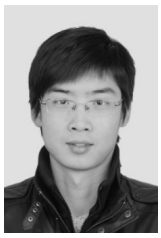
[44] WANG W, YAN Z W. A survey of the DNS cache service in China[J]. Modern Preventive Medicine, 2015.

[45] COX R, MUTHITACHAROEN A, MORRIS R T. Serving DNS using a peer-to-peer lookup service[C]//The International Workshop on Peer-to-Peer Systems. 2002.

[46] MUELLER M L. Competing DNS roots: creative destruction or just plain destruction[C]//Research Conference on Communication, Information. 2001.

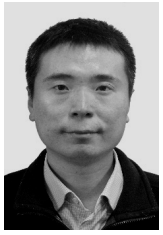
[47] IAB. IAB technical comment on the unique DNS root[S]. IETF RFC2826, 2000.

作者简介：



延志伟（1985-），男，山西兴县人，博士，中国互联网络信息中心副研究员，主要研究方向为 IPv6 移动性管理、BGP 安全机制、信息中心网络架构。

耿光刚（1980-），男，山东泰安人，博士，中国互联网络信息中心研究员，主要研究方向为机器学习、大数据分析和互联网基础资源安全。



李洪涛（1977-），男，河北保定人，中国互联网络信息中心高级工程师，主要研究方向为 IPv6、网络安全、大数据。

李晓东（1976-），男，山东菏泽人，博士，中国互联网络信息中心研究员，主要研究方向为互联网基础资源管理及网络安全技术。