哈爾濱工業大學

毕业设计（论文）开题报告

DNS 根服务数据分析系统设计与实现

专　　业　信息安全

学　　生　朱新锐

学　　号　1180300203

指导教师　詹东阳

日　　期　2021 年 11 月

哈尔滨工业大学教务处制

目录

1. 课题来源及研究的目的和意义

DNS 是域名系统（Domain Name System）的英文缩写，是互联网的一项基本服务。根域名服务器是在域名系统的分层树型结构中处于最顶层的服务器，是支撑万维网运行的最重要的基础设施之一。离开了根域名服务器，DNS 将无法正常运行。[1]

受限于早期 DNS 协议中的限制，全世界只有 13 台 IPv4 根域名服务器，它们的具体分布情况如表 1-1 所示。为了适应日益增长的 DNS 解析需求，一些国家、地区或组织机构引入了根服务器的根镜像服务器，来提供与根服务器相同的功能。中国大陆没有架设 IPv4 根服务器，但部署了 F I J K L 五台根服务器的共 22 台根镜像服务器，它们的具体分布情况如表 1-2 所示。[2]

就 IPv6 根服务器而言，相关的根服务器测试和运营实验项目“雪人计划”已于 2015 年 6 月 23 日正式发布，于全球多个国家架设 25 台根域名服务器。项目将于中国架设一台主根、三台辅根，共 4 个根服务器。

从某种意义上来讲，作为互联网基础设施的 DNS 根服务器是互联网的命脉，如果根域名服务器或其镜像的服务出现故障，导致某些地区的 DNS 解析服务受到影响，乃至于停止服务，那么那些依靠受影响的域名系统所支持的互联网应用和服务也将停止工作。而研究相关 DNS 根服务器、DNS 根镜像的安全问题则需要收集 DNS 根服务器与根镜像服务器的基本信息。从这个角度来看，我们有必要建立一个根服务数据分析系统，用于分析由相关测量系统收集到的有关 DNS 根服务器或 DNS 根镜像的数据。分析这些数据有助于实时感知根服务器的运行状态，可以为类似于根服务器停止服务这样事件的应急响应提供支持。同时收集到的数据也可以用于评估测量节点所在地的 DNS 解析服务水平，为 DNS 根镜像服务器的部署提供参考。

名称	主服务器运营者	主服务器位置	IP
A	INTERNI.NET	美国弗吉尼亚州	198.41.0.4
B	美国信息科学研究所	美国加利弗尼亚州	128.9.0.107
C	PSINet 公司	美国弗吉尼亚州	192.33.4.12
D	马里兰大学	美国马里兰州	128.8.10.90
E	美国航空航天管理局	美国加利弗尼亚州	192.203.230.10
F	因特网软件联盟	美国加利弗尼亚州	192.5.5.241
G	美国国防部网络信息中心	美国弗吉尼亚州	192.112.36.4
H	美国陆军研究所	美国马里兰州	128.63.2.53
I	Autonomica 公司	瑞典斯德哥尔摩	192.36.148.17
J	VerSign 公司	美国弗吉尼亚州	192.58.128.30
K	RIPE NCC	英国伦敦	192.0.14.129
L	IANA	美国弗吉尼亚州	198.32.64.12
M	WIDE Project	日本东京	202.12.27.33

表 1-1 IPv4 根服务器分布表

根镜像类型	管理机构	部署城市	该城市部署的该类型根镜像数量
F	ISC	北京	1
		杭州	1
		西宁	1
		武汉	1
I	Netnod	北京	1
J	Verisign	北京	1
		杭州	1
K	RIPE NCC	北京	1
		广州	1
		贵阳	1
L	ICANN	北京	4
		上海	1
		郑州	2
		武汉	2
		西宁	2
		海口	1

表 1-2 中国大陆 DNS 根镜像分布

2. 国内外在该方向的研究现状及分析

根据测量方式的不同，DNS 测量技术可以分为主动测量与被动测量。

对于被动测量而言，BrownLee 等人[3]于 2001 年针对根服务器中 F.root-server.net.服务器捕获 DNS 数据报文。在对捕获到的数据和服务器日志进行统计分析后，发现 60%~85%的查询请求是来自同一主机的重复的请求，超过 14%的查询请求违反了 DNS 查询的协议。涉及根服务器的拒绝服务攻击也是较常见的。这些都增添了根服务器的负担。

对于主动测量而言，2012 年，杜跃进等人[4]从终端用户角度出发，为 DNS 解析网络性能选取参数，实现了对全国各省市的 60 个 DNS 递归解析服务器进行主动测量，获取了全国 DNS 解析路由路径平均长度，DNS 解析的平均时延等数据。2013 年，孙瑞[5]设计并实现了基于分布式平台的 DNS 信息探测系统，在分布式环境下，采用主动探测的方法实现了 DNS 信息探测，同时还实现了对 DNS 服务器软件版本信息的探测。2016 年，Jones B 等人[6] 利用 RIPE ATLAS 等收集全球网络信息的工具，提出了一种检测未经授权的 DNS 根镜像服务器的技术，并以此方法测定了若干 DNS 代理和一个 DNS 根镜像。Rijswijk-Deij 等人[7] 则实现了对互联网上的主要顶级域名下的域名（如.com、.net、.org）进行长时间、大规模的 DNS 主动测量与 DNS 信息收集，还研究了降低这种行为对 DNS 服务器负担的

方法。2019 年，Callejo P 等人[8]提出了一种新颖而灵活的 DNS 测量方法，利用在线广告的拓展性来分发基于 JavaScript 的轻量级网络测量脚本，并利用该脚本收集数据，来研究递归 DNS 服务器的基本结构、DNS 服务器部署策略和全球用户的 DNS 服务商选择。

近年来 DNS 探测也实现了主被动相结合的方法。王锐[9]于 2021 年实现了主被动探测相结合的 DNS 探测方法，首先使用被动探测技术收集分析 DNS 服务器信息，然后进行扫描式的探测来获取 DNS 服务器信息。

3. 主要研究内容

研究内容主要是设计并实现一个 DNS 根服务数据分析系统，该系统是 DNS 态势感知系统的一部分，用于满足对 DNS 根服务数据测量系统收集到的数据进行初步分析的需求，并为后续的研究提供数据支持。由于许多地区的 DNS 解析服务是由 DNS 根镜像而非 DNS 根域名服务器本身提供，除了 DNS 根域名服务器的相关数据，本系统计划收集的数据也包括 DNS 根镜像的相关数据。

具体内容分为两部分，其一是 DNS 根服务数据分析系统整体的设计与实现，其二是数据分析方法的设计与实现。

对于整体系统的设计与实现，本系统应当具备收集来自 DNS 根服务测量系统的数据、并能根据选定的分析方法实时分析测量数据，同时应当具备格式化存储测量数据与分析数据的能力。如果相协作的 DNS 探测系统开放相关权限的话，本系统还应具备调度分配测量任务的能力，因为在某些情况下可能需要根据分析结果调整测量系统的测量任务。最后系统应当能以图表的形式向操作人员展示数据。该系统应当可以作为 DNS 态势感知系统的一部分，为关于 DNS 的应急响应提供支持。

对于数据分析方法的设计，需要确定从原始数据中收集估计 DNS 服务器信息的方法。设计的方法应当能至少得到关于 DNS 服务器的延迟、可用性的信息，能进行简单的异常检测并响应。

4. 研究方案

研究方案分为两个部分：DNS 根服务数据分析系统的设计与实现，和数据分析方法的设计与实现。

4.1 DNS 根服务数据分析系统的设计与实现

现阶段主要考虑收集处理国内的 DNS 根镜像数据。尽管预期的数据量较小，使用集中式的数据分析系统也能完成分析与存储工作，但主要出于提高可靠性的目的，以及考虑到对测量任务进行分配调度的需求，依然将 DNS 分析系统设计为分布式的系统。分析系统采用分布式的架构将增大系统的冗余度，提高可靠性，单台机器的宕机不至于影响系统运行，同时若以后扩大 DNS 根数据收集的范围，采用分布式架构也方便增强处理能力。另外，因为不同地区享受到的 DNS 解析服务能力并不相同，采集 DNS 根服务数据时有必要

收集不同地区产生的探测数据，因而对应的 DNS 根服务数据测量系统很可能会采用分布式的设计，而这也意味着本系统需要拥有调度测量任务的能力，需要一个任务调度模块。

分析系统具体由任务调度模块、分布式实时计算引擎、数据库、微服务模块和 Web 端组成，系统架构如图 4.1 所示。

任务调度模块负责接收由分布式 DNS 测量系统产生的实时测量数据，并将数据分发给计算引擎。尽管预期的平均流量较小，但考虑到测量节点通常是周期性的返回测量数据，存在短时间内集中返回数据的可能。为了预防突然性的大流量数据影响下游引擎的运行，该模块应当具有将数据流削峰的功能。该模块同时还负责调度 DNS 测量任务，调度任务既包括通过合理安排各个测量节点测量任务的工作量与任务时间来，尽量避免大量测量节点在短时间内集中返回数据，也包括处理根据分析结果调整测量任务的请求。该模块也负责计算引擎的负载平衡工作，降低对计算引擎的性能要求。

分布式实时计算引擎负责用于加工处理接收到的探测数据，并将处理结果交给数据库。这里采用分布式架构的主要目的是提高可靠性，因此只需部署少量计算引擎。

数据库模块存储接收到的原始数据与计算引擎输出的结果，并提供查询接口。由于要存储的数据量预计仅在 GB 级别，因此不计划采用分布式数据库，使用单个数据库以提高查询效率，同时配合数据备份来提高可靠性。

微服务模块位于数据库和用户的中间层，为用户提供查询、安排测量任务的等服务，同时实现对用户的访问控制，保护数据库的安全。该模块将使用 springboot 与 springcloud 技术实现。

最后 Web 端用于提供操作界面与图表展示。该模块将通过 Vue 技术实现，会周期性请求微服务获取数据，然后将数据进行动态展示。

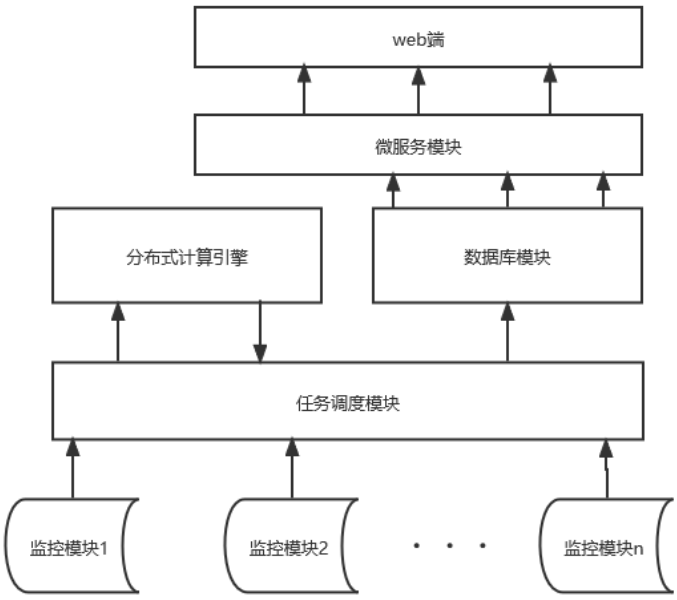

图 4-1 系统框架图

4.2 数据分析方法的设计与实现

需要实现分析 DNS 根服务器及 DNS 根镜像的延迟与可用性，和简单的异常监测。假设 DNS 根服务数据测量系统会周期性提供关于目标根服务器或根镜像的一组测量数据，每组数据包含一个周期内测量节点多次测量得到的数据。

分析 DNS 根服务器或其镜像的延迟的目的是估计测量节点所在地的 DNS 解析服务水平，较小的延迟意味着较好的服务。系统将从测量数据中解析出实时延迟、参考延迟作为输入，计算出延迟指数作为输出。

实时延迟数据来源于 DNS 根服务数据测量系统返回的测量数据，记录某个测量节点在某个时刻与某 DNS 根服务器或根镜像间的查询延迟。考虑到 DNS 根服务器或其镜像在 IPv4 和 IPv6 协议下的服务能力可能有差异，预期的测量系统返回的实时延迟应分为 IPv4 下的 udp 延迟、IPv4 下的 tcp 延迟、IPv6 下的 udp 延迟、IPv6 下的 tcp 延迟四部分，每部分也应在一个探测周期内有多条数据，来减少偶然因素的影响。

参考延迟数据也来源于 DNS 根服务数据测量系统返回的测量数据，记录该地与若干较权威的 DNS 递归服务器间的延迟，目的是为评估实时延迟的高低提供参考。这些 DNS 递归服务器是事先选定的。

延迟指数用于在一定程度上评估实时延迟的相对大小，由当前的实时延迟数据、参考延迟数据与近期的历史实时延迟数据计算得来。目前先假设延迟值符合正态分布，利用近期的历史数据估计分布的参数，然后依次计算当前实时延迟相对理论均值的方差。然后再用这个方差跟当前实时延迟与参考延迟间的比值做积，作为某个测量节点返回的某组数据的延迟指数值。可以考虑设定一个阈值，当指数超过阈值后，系统写入日志并给出提示。

分析 DNS 根服务器或其镜像可用性的目的是估计测量节点所在地获取 DNS 解析的能力。这个功能要求测量系统提供的数据包含一个布尔型的可达性标志，这个标志用于描述在一次测量中测量节点的主动测量数据报是否到达目标服务器。考虑到 DNS 根服务器或其镜像在 IPv4 和 IPv6 协议下的服务能力可能有差异，同样预计这个数据包含 IPv4 下的 udp 可用性、IPv4 下的 tcp 可用性、IPv6 下的 udp 可用性、IPv6 下的 tcp 可用性四部分。运行分析算法的计算模块将以某个测量节点一个周期内针对某台服务器的一组测量数据作为输入，以一个布尔值作为输出，用这个布尔值描述该节点所在地关于该服务器的可用性。计算模块将提取这组数据中的所有可达性标志，若其中表示可达的标志数不少于三个，就输出真值，表示对应测量节点能享受到该服务器提供的 DNS 解析服务。

关于对可用性的简单异常监测，仅仅是出现了某组数据可用性为伪的情况并不会被认为是异常，因为只需与少数几个 DNS 服务器间保持畅通便足以得到堪用的 DNS 解析服务。因此，仅当所有目标根域名服务器和根镜像都对某个测量节点不可用或某个目标根（镜像）服务器对大量测量节点都不可用，才会被当成异常情况处理。此时系统向 DNS 根服务信息测量系统发出临时性的测量请求，要求出现目标 DNS 根服务器或根镜像不可达现象的测量节点立即测量，尝试复现之前不可达的现象，以确认某地 DNS 解析服务确实可能出现问题，或某 DNS 根服务器或根镜像可能出现问题，然后向用户发出警报。

5. 进度安排，预期达到的目标

1. 2021 年 11 月：阅读文献，撰写开题报告。

2. 2021 年 12 月－2022 年 1 月：继续收集、分析资料，形成具体工作思路。

2. 2022 年 3 月- 2022 年 4 月：初步完成分析系统的雏形，进行调试并发现不足之处。

3.2022 年 5 月：改进系统，撰写毕业论文。

6. 课题已具备和所需的条件、经费

1.本人对这个课题有兴趣，前期已经阅读了一定数量的相关文献，经过向指导老师多次请教，已经形成了初步工作思路，下一步就是继续收集、分析相关资料，使思路进一步细化、具体化，为实际完成分析系统打下基础。

2.实验分析方面，学院实验室具备进行相应实验分析的条件。

3.经费方面，目前所有相关资料可以从学校图书馆数字资源中免费获取，不需要经费。

7. 研究过程中可能遇到的困难和问题，解决的措施

1.DNS 根服务测量系统提供的数据中可能会因为故障等原因出现无效数据，拟采取的解决措施为查阅相关资料，明确此类数据的定义并寻找剔除此类数据的方法。

2.实验系统运行时可能出现故障，导致实验数据丢失，拟采取的解决措施为做好实验数据的定期备份工作。

8. 主要参考文献

[1] 于世梁. 浅谈根域名服务器与国家网络信息安全[N].江西行政学院报. 2013 年 4 月，第 15 卷第 2 期

[2] Root Server Technical Operations Association［EB /OL］.2021-11-25． https://root-servers.org/

[3] Brownlee N , Claffy K , Nemeth E . DNS measurements at a root server[C]// IEEE Global Telecommunications Conference. IEEE, 2001.

[4] 杜跃进,张兆心,王克,杨逍,胡萍.基于用户感知的 DNS 解析网络性能测量技术[J].南京航空航天大学学报,2013,45(01):110-115.

[5] 孙瑞. 基于分布式平台的 DNS 信息探测系统设计与实现[D].哈尔滨工业大学,2013.

[6] Jones B , Feamster N , Paxson V , et al. Detecting DNS Root Manipulation[C]// International Conference on Passive and Active Network Measurement. Springer International Publishing, 2016.

[7] Rijswijk-Deij R V , Jonker M , Sperotto A , et al. A High-Performance, Scalable Infrastructure for Large-Scale Active DNS Measurements[J]. IEEE Journal on Selected Areas in Communications, 2016.

[8] Callejo P , Cuevas R , Vallina-Rodriguez N , et al. Measuring the Global Recursive DNS Infrastructure: A View From the Edge[J]. IEEE Access, 2019, 7:1-1.

[9] 王锐. 互联网 DNS 服务资源测绘技术研究[D]，北京邮电大学.2021.