

一种智能 DNS 的设计与实现^①

张立成¹, 彭勇华²

¹(广东药学院 现代教育技术中心, 广州 510006)

²(广州鼎通计算机科技有限公司, 广州 510630)

摘要: 分析了目前包括 BIND 9 在内的各种域名定向解析方法的不足, 提出并实现了一种改进的 SmartDNS 解决方案, 并以实验证明, SmartDNS 是灵活的、有效的, 其解析结果要优于传统的智能 DNS。

关键词: 多链路; 域名服务; 域名解析; 智能 DNS

Design and Implementation of a Smart DNS System

ZHANG Li-Cheng¹, PENG Yong-Hua²

¹(Network Center, GuangDong Pharmaceutical University, Guangzhou 510006, China)

²(Dtone Networks Company, Guangzhou 510630, China)

Abstract: The paper demonstrates the disadvantage of the traditional strategies of domain name directional resolution including BIND9, then we present a preferable system: SmartDNS. By practice, it is available and flexible, and can get better results than traditional means.

Key words: multiple Link; DNS serves; DNS resolution; smart DNS

为提高公网的访问速度及避免高额的教育网国际流量, 许多学校都采用多出口链路, 当校园网用户访问教育网资源时, 就路由到教育网出口, 当访问公网及国外的资源时, 就路由到公网出口进行访问。

1 基于IP的域名定向解析

虽然采用多出口链路, 就解决了校内用户访问互联网所面临的以上两个问题, 但当校外的用户访问学校公共资源时, 仍然会存在一些问题。例如: 学校一般都采用教育网 IP, 作为对外提供服务的服务器 IP 地址(如 Web、DNS 服务等)。当某校外用户访问学校网站时, 若该用户源自教育网, 访问速度一般会较正常, 但若该用户源自公网时, 访问速度可能就会很慢。

为了充分利用公网出口资源, 提高外网用户访问学校网站的速度, 通常会给该网站服务器另配一个公网 IP, 同时也设定一个对应的公网域名。当外网用户访问学校网站时, 就让用户自己选择输入教育网域名还是公网域名, 这是最原始的解决办法。该方法的缺点是不够灵活智能, 完全依靠用户手工选择, 同时多套

域名将会使学校的域名变得不再统一。

我们最理想的实现目标是: 只采用统一的一套域名, 不需用户手工选择, 而是依据访问用户源自哪个网络, 智能的应答给用户相对应的 IP。目前有许多方法可实现该目标。主要的方法有:

1.1 策略路由方法实现

该方法主要是在路由器上进行策略路由设置, 依据访问的源 IP 来进行路由跳转, 当访问者源自教育网时, 就路由到学校教育网 DNS 作解析, 当源自公网时, 就路由到学校的公网 DNS 作解析。这种方法的缺点是配置起来比较烦琐, 维护更新也比较麻烦^[1]。

1.2 通过负载均衡设备来实现

该方法是在负载均衡设备上定义服务器组, 再对服务器组设置相关的地址策略, 最后再依据该地址策略和用户的源 IP 来进行相关的地址解析转换, 当用户源自公网时, 则返回公网 IP, 当源自教育网时, 则返回教育网 IP; 该方法的缺点是费用投入高, 一台具备该功能的负载均衡设备价格不菲, 另外就是配置起来也不太容易^[2]。

^① 收稿时间:2012-03-29;收到修改稿时间:2012-05-05

1.3 购买 CDN 服务商的服务实现

CDN 的全称是 Content Delivery Network, 即内容分发网络, 它的基本思想是将服务器的内容镜像分发到多个地方, 当用户访问时, 服务器进行智能识别判断, 让用户获取离用户最近的服务器, 提高访问速度. 该方法的缺点是购买 CDN 服务价格昂贵, 一般只适合于一些较大的门户网站^[3].

1.4 采用一些具有智能 DNS 功能的软件来实现

目前有一些 DNS 服务器软件具有一定的智能功能, 能依据用户的 DNS 请求的源 IP 地址来进行解析. 最典型的如基于 Unix/Linux 平台的 BIND 9 软件, 类似的还有基于 Windows 平台的 WinmyDNS1. 这种方式最大的优点是投入费用比较低、配置及实现起来也相对比较容易, 所以目前该方式是使用最多的.

BIND 是经典的 Unix/Linux 平台 DNS 软件, 具有稳定、安全、可靠性高、成熟等特点, 其 V9 版本通过引入 View 视图特性, 支持基于源 IP 判断的智能 DNS 解析特性, 使用较为广泛. WindmyDNS 的功能同 BIND 9 类似, 但其成熟度不如 BIND 高, 使用较少.

虽然 BIND 9 能实现智能 DNS 解析, 但仍存在一些不完善的地方, 因此, 我们以 BIND 9 为基础, 设计并实现了一套新的智能 DNS 系统, 它弥补了一些 BIND 9 的不足之处.

2 BIND 9 的不足之处

在 BIND 9 中, 通过 View 语句, 预先定义好了教育网网段、公网网段的访问策略; 依据源 DNS 请求的 IP 地址, 来进行判断, 如果源 IP 是属于教育网网段, 则依据策略, 返回服务器的教育网 IP 地址, 如果是属于公网网段, 则返回服务器的公网 IP 地址.

一般情况下, BIND 9 的这种策略方式会运作得较好, 但在某些特殊情况下, 这种方式会出现一些不足, 下面举例来说明.

如图 1 所示, 当公网用户 A 访问网站 www.gdpu.edu.cn 时, 首先通过 DNS 服务器进行解析, 得到 Web 服务器的公网 IP 地址 219.136.229.45, 然后依据路由通过 Line 1 进行访问 Web 服务器. 然而假如一开始 Line 1 就出现拥塞或故障时, DNS 并不能感知该拥塞或故障的发生, 仍然解析给用户 A 一个公网 IP 地址, 此时, 用户 A 依据此地址进行访问将会有较高的时延或访问失败. 若此时 Line 2 较为顺畅, 显然用户 A 更好的选择是通过 Line 2 来进行访问, 这样将能保

证一个更快速或成功的访问. 此时就需要 DNS 服务器解析给用户 A 的地址是一个教育网 IP 地址, 以达到用户 A 是通过 Line 2 访问到 Web 服务器目的, 本文的 SmartDNS 就是为了实现该目标而设计的.

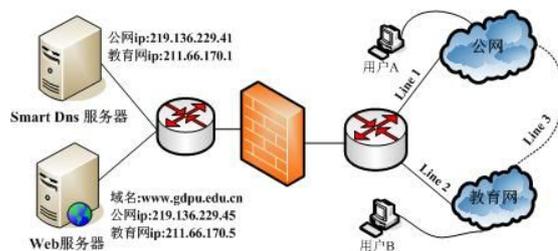


图 1

3 SmartDNS 的设计及实现

与 BIND 9 的 View 不同, 我们的 SmartDNS 的选择算法不仅仅是基于静态 IP 地址来实现的, 它的首选方法是通过动态的网络探测, 来选择适合的服务器 IP, 基于静态 IP 地址段只是 SmartDNS 的一种候选补充方法.

3.1 动态网络探测方法

我们在此称用户的 DNS 服务器为 LDNS(local DNS), 以图 1 为例, 当用户 A 需要访问 www.gdpu.edu.cn 网站时, 需向其 LDNS 服务器提出解析请求, 最终我们的 SmartDNS 将会收到该 LDNS 的请求报文. 在 SmartDNS 内部, 有一张大的 Hash 映射表作为 LDNS 邻近表, 该表的 Key 是各个 LDNS 的 IP 地址, Value 是本地服务器的 IP 地址和过期时间. 刚开始运行时, 该表是全空的. 当 SmartDNS 收到用户 A 的 LDNS 的请求时, SmartDNS 就在该 Hash 表内进行查询, 如果有匹配的 Key, 则返回该 Key 对应的 Value 值 (服务器 IP 地址), 否则就进行网络报文探测.

具体的探测方法是: SmartDNS 分别从公网和教育网两条链路对用户 A 的 LDNS 发送链路探测报文, 从而获得 LDNS 的两条链路的 RTT (数据往返时间) 值, 并对两条链路的 RTT 值进行比较, 返回较小 RTT 值链路所对应的 Web 服务器 IP, 并将该 LDNS 的 IP、Web 服务器 IP 存入 Hash 表, 当下次再收到该 LDNS 的请求时, 就直接从该 Hash 表中获取数据, 无需再进行报文探测, 以提高效率. 由于网络链路状况不是一成不变的, 需对每个 Hash 表项增加一个过期时间值, 当达到过期时间后, 就将该表项从 Hash 表中删除. 为保证数据的准确性, 同时也要考虑到 SmartDNS 的工作效率, 该过期时间值不宜设置得太大, 也不能设置得太

小, 需取一个相对平衡的值. 根据短时间内网络状况变化不大的情况以及多次的实验结果, 我们在此将过期时间值设定为大约 5 分钟.

3.1.1 网络链路探测报文

为获取 LDNS 的 RTT 时间, 需向 LDNS 发送链路探测报文, SmartDNS 可选择多种探测方法进行探测, 主要包括: (1) DNS_Dot: 向 LDNS 发起一个包含“.”的测试, 也就是向目标 DNS 请求 root 清单, 该解析一般默认配置的 DNS 服务器均提供支持^[4]; (2) DNS_REV: 向 LDNS 发起本机 IP 的 PTR 请求; (3) UDP 包: 向 LDNS 发起一个 UDP 的包, 看是否回应; (4) TCP 包: 向 LDNS 发起一个 TCP 的包看是否回应; (5) ICMP: 向 LDNS 发起一个 ICMP 的 ping 包, 看是否回应; 在以上各检测方式中, 无论目标系统返回哪种类型的数据包, SmartDNS 均可认为是有效数据包, 从而记录数据包往返时间, 并进行比较, 最终生成 LDNS 的 Hash 邻近表.

3.2 静态 IP 地址段方法

当从各出口链路发出的探测报文均失效时(没收到探测报文回应包或超时), 就启用静态 IP 地址段方法, 同 BIND 9 的 View 类似, SmartDNS 上可配置多个自定义区域, 并将这些区域与链路相对应. 当用户的 LDNS 发起请求连接 SmartDNS 时, SmartDNS 将根据 LDNS 所位于的区域返回给 LDNS 适当链路服务器 IP 地址, 引导用户从正确的线路进行访问. 在该方法下, 需收集各运营商的 IP 地址网段, 将网段进行整理导入 SmartDNS, 形成自定义区域表.

3.3 健康检查进程

为获得较高的效率和准确率, SmartDNS 引入了一个健康检查进程. 该进程定时的向各网络出口链路发送 ICMP 包, 以检查各出口链路是否正常. 当某链路的丢包率超过一定阈值时, 则认为该链路发生了故障, 从而排除对该链路的动态探测, 以及该链路对应的区域 IP 的结果反馈; 当该出口链路的健康检查包丢包率降到正常水平时, 才将该链路置为正常状态. 在本文的例子中, (即只有两个出口链路的情况), 当检查到有一个链路出现故障时, 则停止进行网络探测行为, 以单出口链路的模式进行运作, 这时同传统的 DNS 运作方式完全一样.

3.4 算法

SmartDNS 的 DNS 解析算法如下:

- (1) SmartDNS 等待接收 DNS 请求;
- (2) 接收到 LDNS 的请求, 判断 $|L-L'| \geq 2$, 若是则转到 step (4);
- (3) Reply IP_i (i 为唯一正常线路) 并转到 step(1);
- (4) 在 Hash 表中查找 Key 值 IPLDNS, 若查找失败, 则转到 step(9);
- (5) 得到表项 (IP_{LDNS}, IP_j, T) , 判断 T 是否超过老化时间, 若是则转到 step(8);
- (6) 如果 $I_j \in L'$, 则转到 step(8);
- (7) Reply IP_j , 并转到 step(1);
- (8) 删除 Hash 表项: Delete (IP_{LDNS}, IP_j, T) ;
- (9) 进行网络探测 Prob $(IP_{LDNS}, L-L')$, 如果失败, 则转到 step(11);
- (10) 得到较优线路 I_k , Reply IP_k , 将该结果添加到 Hash 表中: Insert (IP_{LDNS}, IP_k, T) , 并转到 step(1);
- (11) 以静态 IP 段方法进行回应, 即
 - If $IP_{LDNS} \in Zone_i$ and $I_i \notin L'$
 - Reply IP_i , 并转到 step (1)
 - Else Reply 候选 IP_m , $I_m \notin L'$, 并转到 step(1)

说明: L 为所有出口链路集合, 设出口链路共有 n 条, 则 $L = \{I_1, I_2, I_3, \dots, I_n\}$, 其中 $1 \leq i, j, k, m \leq n$, L' 为非正常线路集合, T 为 Hash 表项的生成时间, LDNS 为发送请求的 DNS 服务器, IP_{LDNS} 为 LDNS 服务器的 IP 地址; 整个算法处理过程不能超过 3 秒, 因为 3 秒是 DNS 查询的 Timeout 时间.

4 主要优点

如上所述, SmartDNS 的最大优点在于能动态、实时的感知网络的状况, 依据当前的网络状况, 返回对用户来说网速最快的资源服务器 IP 地址.

除此之外, 当存在多台相互冗余的资源服务器时, SmartDNS 还能实现多台服务器的负载均衡. SmartDNS 的健康检查器除了进行链路状况的检测外, 还可对服务器的各项指标进行检测分析, 包括服务器响应时间、CPU 利用率、连接比率、任务完成率、最少连接数等, 通过在服务器上安装的健康检查代理进程, 定期的收集服务器的各项指标, 计算出一个综合负载值, 传递给 SmartDNS 的检查进程, SmartDNS 依据各服务器的状况生成相应的权值, 以此权值生成服务器的轮询队列, 从而实现服务器的负载均衡.

另外 SmartDNS 还能进行流量定向功能. 由于教

育网的国际流量收费较高,因而教育网的国际出口一般都会较顺畅,当 SmartDNS 对源自国外的 LDNS 请求进行 prob 时,得出的最快链路很可能是教育网链路,为避免高额的国际流量费用,可进行流量定向,排除教育网链路,即当收到源自国外的 LDNS 请求时,不进行教育网的链路探测,这样最终返回给用户的服务器资源 IP 也不会是教育网的 IP 地址,从而避免了经过教育网的流量,节省了网络带宽费用。

5 SmartDNS的测试

为了验证 SmartDNS 的正确性和可靠性,我们选择了 nslookup 和 ping 两个工具来进行测试。

本校的校园网有三个网络出口链路,网站服务器有三个互联网 IP 地址,分别是教育网 IP1(211.66.170.5)、电信网 IP2(219.136.229.45)、联通网 IP3(210.21.94.135),我们分别用源自三个网络的客户机来对网站服务器的域名 www.gdpu.edu.cn 进行解析,测试经 SmartDNS 解析出来的 IP 是否是最优的。

5.1 正常状况下的实验

各链路出口正常,对域名进行 20 次的解析结果:

测试客户机	IP1	IP2	IP3
Client 教育网	20	0	0
Client 电信网	0	20	0
Client 联通网	0	0	20

三个客户机对三个 IP 地址的 ping 值平均时延:

测试客户机	IP1	IP2	IP3
Client 教育网	5.8	21.4	43.5
Client 电信网	20.1	11.2	33.4
Client 联通网	45.2	36.7	18.6

从以上数据可看出,在出口正常情况下,SmartDNS 均能返回给用户最优的解析结果。

为测试 SmartDNS 的有效性和灵活性,我们使用流控设备将任意出口调节为拥塞状态,再进行观察。

5.2 出口非正常情况下的实验

5.2.1 教育网拥塞,电信网和联通网正常的实验

对域名进行 20 次 nslookup 的解析结果:

客户机	IP1	IP2	IP3
Client 教育网	1	19	0
Client 电信网	0	20	0
Client 联通网	0	0	20

三个客户机对三个 IP 地址的 ping 值平均时延:

测试客户机	IP1	IP2	IP3
Client 教育网	768.3	27.6	44.5
Client 电信网	910.4	14.2	34.1
Client 联通网	1045.6	33.3	17.4

5.2.2 电信网拥塞,教育网和联通网正常的实验

对域名进行 20 次 nslookup 的解析结果:

测试客户机	IP1	IP2	IP3
Client 教育网	20	0	0
Client 电信网	13	0	7
Client 联通网	0	0	20

三个客户机对三个 IP 地址的 ping 值平均时延

测试客户机	IP1	IP2	IP3
Client 教育网	6.3	1123.5	41.5
Client 电信网	23.1	1020.3	38.6
Client 联通网	47.2	1108.4	15.4

由以上实验数据可看出,在出口线路出现拥塞或中断的情况下,SmartDNS 均能灵活的进行动态调整,最终返回给用户最优的结果。

6 结论

本文针对传统的 DNS 系统进行分析,指出了传统的 DNS 所存在的缺点及不足,提出并实现了一套改进的 DNS 系统 SmartDNS,并且又同功能较接近的 BIND 9 进行了比较分析,说明 SmartDNS 要比 BIND 9 的适应性更强、更灵活。最后通过实验进行测试,在各种网络链路条件下,SmartDNS 均能返回最优的结果,进一步说明 SmartDNS 是有效的、灵活的。

参考文献

- 1 许东民,任宇.基于智能 DNS 的校园网络双出口的设计和实现.成都信息工程学院学报,2007,22(6):716-718.
- 2 葛昕,岳敏楠.校园双出口 DNS 负载均衡的应用研究.计算机系统应用,2008,17(9):105-107.
- 3 李文正,郭巧,王利,郭为民.Internet 服务器负载均衡的研究与实现.计算机工程,2005,31(6):98-99,165.
- 4 Mockapetris P. Domain Names-Concepts and Facilities (RFC1034). IETF. www.ietf.org/rfc/rfc1034.txt.