计算机科学 2002Vol. 29№. 8

DNS:域名系统分析与研究

Analyse and Study for DNS: Domain Name System

钟乐海

(四川师范学院计算机科学系　南充 637002)

Abstract Domain name system based on the user datagram protocol's is one of the most important parts of the TCP/IP protocols family of Internet. It affects all Clients and Servers' smooth communication on the whole net. Analyzing and researching of the Domain Name System's framework structure and working ways can also make us understand the Internet's concrete structure and know the TCP/IP protocols much deeper. We unclose its veil after analyzed Domain Name System's datagram patterns, pointer inquiring, resource record and took examples.

Keywords Domain name system, DNS datagram, Pointer inquiring, Resource record, Cache, Unix operation system

1 引言

域名系统(DNS)是一种用于 TCP/IP 应用程序的分布式数据库,它提供主机名字和 IP 地址间的相互转换及有关电子邮件的选路信息。换句话说 DNS 提供了允许服务器和客户程序相互通信的协议。

首先从应用程序角度看,对 DNS 的访问是由一个地址解析器(Resolver)完成的。在 UNIX 主机中,解析器主要是通过调用库函数 gethostbyname 和 gethostbyaddr 来访问的。在编译过后它们与应用程序连接在一起。前者通过主机名得到 IP 地址,后者则通过 IP 地址来找主机名。解析器通过一个或多个名字服务器来完成该转换。但应分清解析器扮演的是工具角色,UNIX 内核中的 TCP/IP 协议族对于 DNS 一点都不知道,是通过解析器将一个主机名转换为一个 IP 地址,从而进行通讯的。

2 宏观的认识 DNS

DNS 的名字空间和 UNIX 的文件系统相似,也具有层次结构。图 1 显示了它的组织形式。每个结点有至多 63 个字符长的标识。树的树根是没有任何标识的特殊结点。命名标识中不区分大小写。命名树上任何一结点的域名就是将从该结点到最高层的域名串起来,中间以"·"分隔,应注意的是这和 UNIX 文件系统路径不同,UNIX 是由树根依次向下形成的。最后,每个结点必须有一个唯一的域名,但不同结点可使用相同标识。

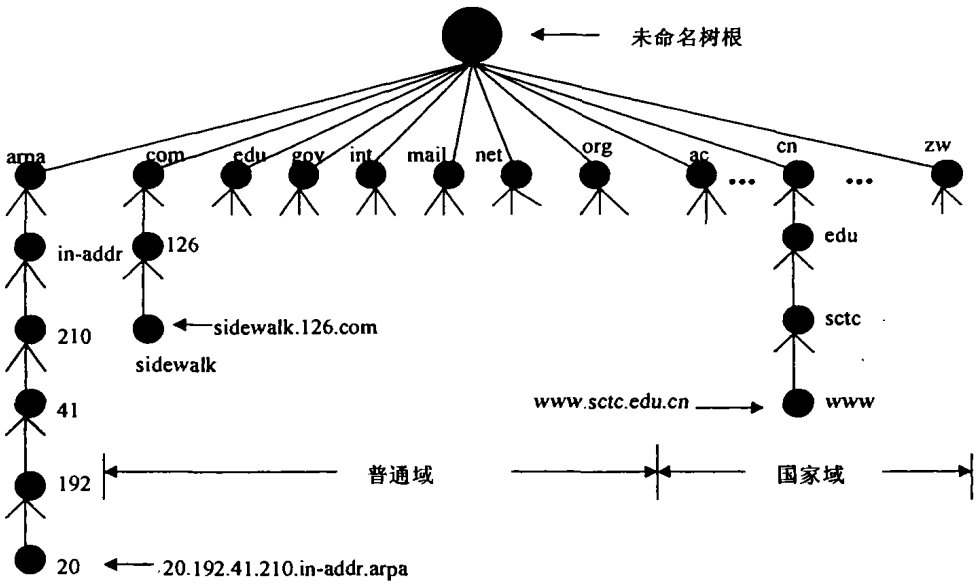

图 1　DNS 的层次组织

以点"·"结尾的域名称为绝对域名或完全合格域名 FQDN(Full Quanlified Domain Name),如 www. sctc. edu. cn。若一个域名不以点结尾,则认为该域名是不完全的。

顶级域名分为三部分:(1)arpa 是一个用作地址到名字转换的特殊域;(2)7 个 3 字符长的普通域,或叫组织域;(3)所有 2 字符长的域均是基于 ISO3166 中定义的国家代码,被称为国家域。

图 2 列出了 7 个普通域的划分,要特别指出的是,在

钟乐海　博士生,副教授,主要研究领域为计算机网络及软件工程。

DNS 中，3 字符长的普通域原本用于美国的组织机构，但现在只有. gov 和. mil 域才局限于美国。2 字符长的国家域用于每个国家。许多国家将它们的二级域也组织成普通域的结构，如：. edu. cn 是我国教育机构的二级域名，. com. cn 是我国商业机构的二级域名。

域	描述
com	商业组织
edu	教育机构
gov	美国政府部门
int	国际组织
mil	美国军事站点
net	网络
org	其它组织

图 2　三字符普通域

一旦一个区域的授权机构被委派后，由它负责向该区域提供多个名字服务器。一个名字服务器负责一个或多个区域。一个区域应获得一个主名字服务器和至少一个辅助名字服务器。主、辅名字服务器必须是独立和冗余的，以便某个服务器发生故障时不会影响到该区域的名字服务。其中主名字服务器是从磁盘中调入该区域的所有信息。区别于它，辅服务器则从主服务器调入所有信息，我们把辅服务器从主服务器调入信息的过程称为区域传送。

在实际应用中，并非每个名字服务器都知道如何与其它名字服务器联系。但是，每个名字服务器必须知道如何同根名字服务器联系。据资料显示到 2000 年为止 Internet 上共有 13 个根名字服务器，其中 10 个在美国，另外英国、瑞典、日本各一个。所有主名字服务器都必须知道根名字服务器的 IP 地址。根服务器则知道所有二级域中每个授权名字服务器的名字和 IP 地址。这意味着一个反复的过程：正在处理请求的名字服务器与根服务器联系，根服务器告诉它去与另一个名字服务器联系。

3　DNS 的报文格式

DNS 定义了一个用于查询和响应的报文格式。图 3 显示了这个报文的总体格式，这个报文由 12 字节长的首部和 4 个长度可变的字段组成。

标题	标志
问题数	资源记录
授权自愿记录数	额外资源记录数
查询问题	
回答(资源记录数可变)	
授权(资源记录数可变)	
额外信息(资源记录数可变)	

图 3　DNS 查询和响应的一般格式

标识字段由客户程序设置并由服务器返回结果，客户程序通过它来确定响应与查询是否匹配。16bit 标志字段划分为若干子字段，如图 4：

QR	OPCODE	AA	TC	RD	RA	(ZERO)	RCODE

图 4　DNS 报文首部标志字段

QR—1bit，0 表查询报文。

OPCODE—4bit，0 表示标准查询，1 为反向查询，2 为服务器状态请求。

AA—1bit，表示授权回答(authoritative answer)

TC—1bit，表示“可截断的(truncated)”。使用 UDP 时，若应答总长超过 512 字节，只返回前 512 字节。

RD—1bit，表示“期望递归(recursion desired)”。此标志告知服务器必须处理这个查询。若该位为 0，且被请求的名字服务器设有一个授权回答，它就返回一个能解答该查询的其它名字服务器列表。

RA—1bit，表示“可用递归”。若名字服务器支持递归查询，则在响应中将该比特设置为 1。随后的 3 比特字段必须为 0。

RCODE—4bit 返回码字段。值为 0 没有差错，为 3 是名字差错，表示查询中制定的域名不存在。

图 3 中随后 4 个 16bit 字段说明最后 4 个变长字段中包含的条目数。查询报文中问题(question)数通常是 1，其它 3 项均为 0。DNS 查询报文中的问题部分，每个问题的格式如图 5。

查询名	
查询类型	查询类

图 5　DNS 查询报文中问题部分的格式

查询名是要查找的名字，它是一个或多个标识符的序列。每个标识符以首字节的计数值来说明随后的标识符的字节长度，每个名字以最后字节为 0 结束，长度为 0 的标识符是根标识符。

计数字节的值必须是 0～63 的数，因为标识符的最大长度仅为 63。图 6 显示了如何存储域名 www. sctc. edu. cn。

3	w	w	w	4	S	c	t	c	3	e	d	u	2	c	n	0

图 6　域名 www. sctc. edu. cn

查询类型—常用是 A 类型，表示期望获得查询名的 IP 地址。一个 PTR 查询是请求一个 IP 地址对应的域名。

查询类—通常是 1，指互联网地址。

回答，授权和附加信息字段均采用一种称为资源记录 RR(Resource Record)的相同格式，如图 7。

类型	类
生存时间	
资源数据长度	
资源数据	

图 7　DNS 资源数据记录格式

域名—是记录中资源数据对应的名字。格式与查询名字段格式(图 6)相同。

类型—说明 RR 的类型码。值与查询类型一样。

类—通常为 1，指 Internet 数据。

生存时间字段是客户程序保留该资源记录的秒数，一般为两天。

资源数据长度—说明资源数据的数量。该数据格式依赖于类型字段的值。对于类型 1，资源数据是 4 字节的 IP 地址。

4　指针查询

DNS 中较难于理解的部分是指针查询方式，即给定一个

IP 地址，返回与该地址对应的域名。首先看图 1，观察一下顶级域名 arpa，及下面的 in-addr 域。当一个组织加入 Internet 并获得 DNS 域名空间的授权，如 sctc. edu. cn，则它们也获得了对应 IP 地址的 in-addr. arpa 域名空间的授权。在 sctc. edu. cn 这个例子中，它是网络号为 210. 41. 192 的 C 类网络。在 DNS 树中接点 in-addr. arpa 的下一级必须是该 IP 地址的第一字节(即为 210)，再下一级为该 IP 地址的下一字节(41)，再后来即为 192，依此类推。同时在此应牢记 DNS 名字是由 DNS 树的底部逐步向上书写的，也就是说 IP 为 210. 41. 192. 20 的 Solaris 主机，其 DNS 名字为 20. 192. 41. 210. in-addr. arpa。同样 FQDN 也是自底向上书写，如果 FQDN 由顶向下写，则这个 IP 地址 DNS 名字将是 arpa. in-addr. 210. 41. 192. 20，而它所对应的域名会是 cn. edu. sctc. www。虽然反写 IP 地址和特殊的域名会造成一些不便，但是要知道如果不这样做就只好从每个树根开始匹配直到顶级域。而现今 Internet 的巨大规模，要完成一个这样匹配也许需要数天或更长的时间。所以比较而言 in-addr 方式是一种最有效的途径。

5 资源记录

在 DNS 的报文中，包含着各种不同的资源记录(RR)。到现在约有 20 种左右，下面将介绍一些很常用的。另外，随着时间的推移，必然还会加入更多的类型。

A—A 记录定义了一个 IP 地址，存储 32bit 的二进制数，即 IP 地址查询。

PTR—用于指针查询。IP 地址被看作是 in-addr. arpa 域下的一个域名。

CNAME—规范名字(canonical name)。用来表示一个域名，而有规范名字的域名通常称为别名(alias)。系统可以用它向其它系统提供一个易记的别名。

假设可以从 ftp. sctc. edu. cn 处免费下载软件，但事实上并没有叫 ftp 的，系统其仅仅是为另一域名相对难记的系统提供的别名。可从 Solaris 主机上用 host 命令查看：

```
Solaris % host -t cname ftp. sctc. edu. cn
ftp. sctc. edu. cn CNAME download. cs. sctc. edu. cn
```

HINFO—表主机信息，包括说明主机 CPU 和操作系统的两个字符串。例如：

```
Solaris % host -t hinfo Solaris
www. sctc. edu. cn HINFO * * * SunOS 5. 7
```

MX—邮件交换记录。许多不能与 Internet 连接的站点通过 UUCP 链路与一个连在 Internet 上的站点相连接。通过 MX 记录可用后者地址向前者发送 E-mail。

```
Solaris % host -t mx sctc. edu. cn
sctc. edu. cn MX mail. sctc. edu. cn
```

NS—名字服务器记录。其说明一个域的授权名字服务器。由域名表示。

```
Solaris % host -t ns sctc. edu. cn
sctc. edu. cn NS www. sctc. edu. cn
```

6 高速缓存等问题

为了减少 Internet 上 DNS 的通信量，所有的名字服务器均使用高速缓存。在标准的 UNIX 实现中，高速缓存是由名字服务器而不是由名字解析器维护的。既然名字解析器作为每个应用的一部分，而应用又不可能总处于工作状态，因此将高速缓存放在只要系统(名字服务器)处于工作状态就能起作用的程序中显得很重要。这样任何一个使用名字服务器的应用均可获得高速缓存。在该站点使用这个名字服务器的任何其它主机也能共享服务器的高速缓存。

另外，我们应注意到 DNS 名字服务器使用的端口号无论对 UDP 还是 TCP 都是 53。那么这意味着 DNS 均支持 UDP 和 TCP 访问，但是正如我们前面说的 DNS 是基于 UDP 的。这又是怎么回事呢？原来，当名字解析器发出一个查询请求，并且返回响应中的 TC(删减标志，见图 4)比特被设为 1 时，即意味着响应长度超过了 512 个字节，从而仅返回前 512 个字节。此时，名字解析服务器通常使用 TCP 重发原来的查询请求，也就允许响应超过 512 个字节。实际之中，当一个辅域名字服务器在启动时，将从该域的主名字服务器执行区域传送。而且运行过程中，辅助服务器也会定时(通常是 3 小时)向主服务器进行查询以便了解主服务器数据是否发生变动。如果有变动，将执行一次区域传送。区域传送就是使用 TCP 的，因为这里传送的数据远比一个查询或响应多得多。所以，可以看出 DNS 是基于 UDP 的说法是仅针对客户端来说的。

既然我们访问 DNS 主要使用的是 UDP，而且查询和响应主要经过广域网，分组丢失率和往返时间的不确定性在广域网上又比局域网更大，所以对于 DNS 客户程序，一个好的重传和超时程序往往显得十分的重要。

总结 综上所述，我们可以看出 DNS 是任何与 Internet 相连的主机必不可少的一部分。通过分析其组织结构、报文格式和各种优化措施，我们可以得出结论：DNS 是一种高效、准确、安全的域名解决方案，它在当今 Internet 上占据着难以替代的地位。

参 考 文 献

1 www. ICANN. org: "Domain Names: Implementation and Specification" "Domain Names: Concepts and Facilities" "DNS Root Nameservers 2000 (july 19 1999)"

2 Internetworking with TCP/IP: Vol. I: The Protocols. Comer, D. E. Prentice-Hall, Eaglewood Cliffs, N. J.

3 TCP/IP Illustrated, Volume 1: The Protocols. W. Richard Stevens. Addison Wesley Longman, Inc.

4 Understanting Data Communications & Networks, Second Edition. William A. Shay. Brooks/Cole Publishing Company