JSERNET UDP53 端口流量分析

苏艳珺，丁伟，方强

（东南大学计算机科学与工程学院，南京，210000）

摘　要：在大多数的防火墙和网络中,53 端口作为 DNS 服务的公认端口，一般不会被屏蔽。这种机制导致其它一些想要穿越防火墙的报文滥用此端口。为全面了解这种现象，本文设计了 UDP53 端口报文检测程序。利用此程序对 JSERNET 的 53 端口进行了流量分析，得到了网络中 DNS 流量及其他非 DNS 流量的比例，还原了网络流的真实情况。

关键词：流量分析；域名系统；　识别

Traffic Analysis of JSERNET's UDP53 Port

Su Yanjun, Ding Wei, Fang Qiang

(School of Computer Science of Engineering, Southeast University, Nanjing, 210000)

Abstract: As the well known port for DNS service, the 53 port won't be shielded by most of firewalls or network services. This defect can be used to send packets through firewalls without being detected. For comprehensive understanding of this phenomenon, a program was designed to inspect UDP flows going through the 53 port. With the help of this program, we conducted traffic analysis on the 53 port of JSERNET and obtained the proportions of DNS flow and non-DNS flow, which were used to describe the real components of such flows.

Key words: Traffic analysis; Domain Name System; Identification

因特网发展初期，公用端口被紧密绑定于一些服务，如 HTTP 流量使用端口号 80，DNS 流量使用端口号 53。所以在设计安全机制的时候，防火墙一般会屏蔽那些端口号不能被识别的流量，而公用端口则会被设置成开放的状态。然而随着互联网规模的不断扩大和各类网络应用的兴起，一些应用或者攻击常常使用这些公用端口号来伪装自己，利用防火墙的这一特征来穿越路由器。因此，传统的基于端口的网络协议判定方法在如今的网络中受到了阻碍。

域名系统（DNS）是互联网的核心服务之一，它使得互联网用户可以使用更加人性化的域名来对网络节点进行标识，而无需记忆数量庞大的 IP 地址，从而为其访问某个网络提供了方便[1]。53 端口是 DNS 服务使用的端口，客户端向 DNS 服务器的 53 端口发出查询请求，服务器解析后将结果返回给客户端。但是现在的网络中，使用 53 端口的除了 DNS 报文外 ，还有很大一部分的穿越流量。所以为了了解网络中真正 DNS 报文的流量状态，我们对 JSERNET 的 UDP53 端口进行了流量分析，得到网络中 DNS 流量的比例与其他非 DNS 流量的比例。

本实验主要利用在 CERNET 华东（北）地区网络中心江苏省网边界采集得到的 IP Trace，根据 IP Trace 的格式设计了 UDP53 端口报文检测程序，并且利用制作的标准答案，确定了此程序对 DNS 报文的查准率。本文从实测数据中观察到 2005 年的 IP Trace 中 DNS 报文占 UDP53 端口报文 99%以上，而到 2009 年后 DNS 报文只占 UDP53 端口报文的 10%左右或者更低，此现象充分表明了对 53 端口的非正常使用越来越多，通过对这些问题 IP 的分析，确认了其中某些 IP 使用 53 端口的主要目的及用途。

1 研究背景

1.1 IP Trace

本论文的研究工作基于 CERNET 华东(北)地

作者简介：苏艳珺，（1988-），女，硕士研究生，E-mail: yjsu@njnet.edu.cn；丁伟，（1962-）女，教授，博导，E-mail: wding@njnet.edu.cn；方强，（1987-）男，硕士研究生，E-mail：qfang@njnet.edu.cn

区网络中心，为了支持网络数据的整理和分析而采集的 IP TRACE。IP Trace 的采集点构建在江苏省网边界。在实验中所用到的数据，会按图 1 的格式进行存储，以 200M 大小为单位组织成一个二进制文件。作为以研究为目的且是设在主干网上的采集系统，考虑存储的有限性和其他网络用户的隐私，采集器采集数据的长度需要有限制，因此全报文的采集方式不被支持。本实验的主要难点就是如何利用这些不完整的报文来筛选出真正的 DNS 报文。

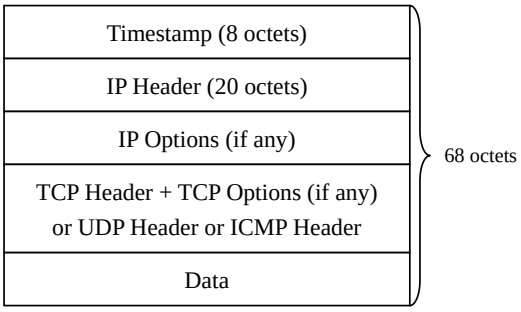

图 1 每个数据包的具体格式

1.2 研究现状

由于越来越多的新兴网络服务选择 UDP 作为其底层的传输协议。因此对于 UDP 流量的研究在近几年也受到学者们的关注。其中不乏对 UDP 端口使用情况的统计。张艺濒等人在文章中就列举了国内某骨干路由器上端口的使用情况，其中 53 端口的使用频率高居第二位。同时在文章的另一张表格，我们可以看到 53 端口上的流数占总流数的 8.6%。而在 CAIDA2009 年的一篇报告中，2009 年 53 端口上流数占总流数的比例只有 3%-4%。这两个数据间的差异让我们对这些 53 端口的流量产生了质疑。

但是对于 UDP53 端口的研究主要围绕在 DNS 攻击这一方面。DNS 是大部分网络的基础，但是由于协议本身的设计缺陷，没有提供信息保护和认证机制，使得 DNS 很容易受到攻击[2]。所以一直以来学者都在探讨 DNS 安全性的问题。而本文只是对 UDP53 端口的流量进行了一个统计，还原了网络流的真实情况，同时也间接解释了上一段提出的质疑。

2 DNS 报文检测

2.1 DNS 报文格式

DNS 定义了一个用于查询和响应的报文格式。表 2.1 显示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[3]。其中标识字段由客户端设置并由服务器返回结果，

客户程序通过它来确定响应报文与查询是否匹配。接着是 16 位标志字段，剩下的 4 个 2 字节分别表示问题数，资源记录数，授权资源记录数，额外资源记录数。4 个长度可变的字段分别对应问题，资源等的具体内容。

表 2.1 DNS 报文格式

0 15	16 31
标识 ID	标志
问题数	资源记录数
授权资源记录数	额外资源记录数
查询问题	
回答	
授权信息	
额外信息	

16 位的标志字段又被划分为如图 2.2 的子字段。其中 QR(1 比特)：查询/响应的标志位，1 为响应，0 为查询。opcode（4 比特）：定义查询或响应的类型。AA（1 比特）：授权回答的标志位。该位在响应报文中有效，1 表示名字服务器是权限服务器 。TC（1 比特）：截断标志位，1 表示响应已超过 512 字节并已被截断。RD（1 比特）：该位为 1 表示客户端希望得到递归回答。RA（1 比特）：只能在响应报文中置为 1，表示可以得到递归响应。zero（3 比特）：保留字段，应该为零。rcode（4 比特）：返回码，表示响应的差错状态。图 2.3 显示的是一个正常的 DNS 应答报文在报文分析软件中的格式，它的标志为 0x8180，问题数为 1，回答数为 3，授权资源记录数为 2，额外资源记录数为 3。

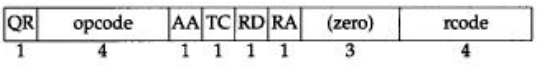

图 2.2 DNS 报文首部中的标志字段

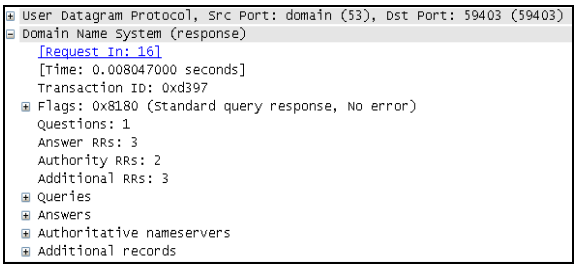

图 2.3DNS 应答报文

2.2 DNS 报文检测原理

由于 IP Trace 存储大小为 68 字节，若 DNS 报文是负载在 UDP 报文上的，除去 8 个字节的时戳，20 个字节的 IP 头部（没有选项）以及 8 个字节的 UDP 头部，余下的 32 字节就是属于 DNS 报文的部分。对报文检测的原理就是利用 12 个字节的 DNS 报文首部必须满足的条件，以下就是本文设计的 3 个条件。

（1）opcode 和 rcode 的最高位必须是零。从图 2.2 看出，opcode 和 rcode 都是一个四位的标志。opcode 是查询类型，例如 0 是标准查询，1 是反向查询，rcode 是服务器设置的返回码字段，例如 0 是没有差错，3 是名字差错。根据目前实行的 DNS 报文协议，opcode 和 rcode 的 0~5 均已被使用，而 6~15 暂为保留，利用这一特性设置了判断条件一。

（2）问题数必须为 1 或者是零。对于查询报文来说问题数必须是 1，而对于应答报文问题数可以是 0 或者 1。

（3）记录数的合理性。根据 DNS 报文的定义，资源记录数，授权资源记录数以及额外资源记录数分别代表了报文体中 3 个字段中记录的个数。同时根据 UDP 首部能够计算得到这三条记录对应的字节数。条件三利用的就是字节数与记录数之间关系。

DNS 报文最后的三个字段是长度可变的字段，它们均采用一种称为资源记录 RR（Resource Record）的相同格式，格式如表 2.4 所示。观察表 2.3，得知 NAME 最小值为 2 个字节（即使用压缩方式，16 位的指针），响应类型，响应类，生存时间，数据长度为定长，总计为 10 个字节，由于资源数据部分存在很大的差异性，所以不能确定其最小值，选择省略。综上，本文设定一个 DNS 资源记录最小的长度为 12 个字节。

表 2.4 DNS 资源记录格式

0	15	16	31
NAME（不定长）			
响应类型		响应类	
生存时间			
数据长度		资源数据（不定长）	

设三条记录数对应的字节数为 rss_len，rss_len 的大小需要将从 UDP 首部中得到的 UDP 长度减去 8 个字节的 UDP 首部和 12 个字节的 DNS 首部以及问题的长度。问题的长度也是一个非定长的，问题的格式如表 2.5 所示，由于查询名不定且无规律，所以将其忽略，只考虑查询类型以及查询类的长度，为 4 个字节。这样最后得到的 rss_len 即为最大的 rss_len。

表 2.5 DNS 报文中问题部分的格式

0	15	16	31
查询名（长度可变）			
查询类型		查询类	

如果用最大的 rss_len 去除以最小的记录长度，算得的就是记录数的上限。若在首部中得到的记录数之和大于上限，就表现出不合理性，也就可以判断这个报文一定不是 DNS 报文。

2.3 报文举例

图 2.6 显示的是一个正常的 DNS 查询报文在报文分析软件中的格式。此报文的 opcode 为 0，rcode 为 0，满足条件一；Questions 为 1，满足条件二；此报文对应的 UDP 数据长度为 39(图中未标出)，则对应的 rss_len 为 15，算得记录数上限为 1，而图中记录的和为 0，满足了条件三，因此这种报文将会被判断为真正的 DNS 报文，与实际情况相符合。

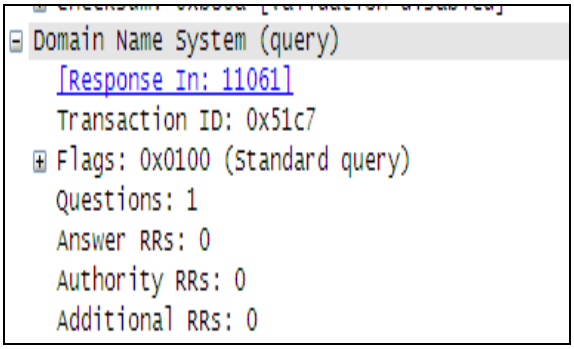

图 2.6 正常 DNS 报文

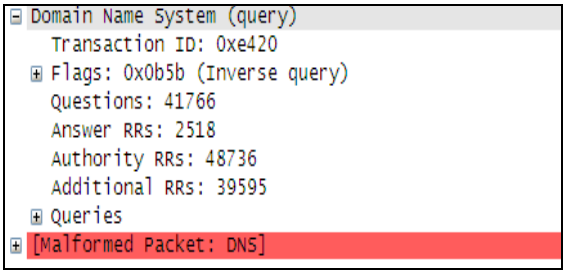

图 2.7 穿越报文

图 2.7 显示的是一个穿越报文在分析软件中的格式，code 为 11，不满足条件一，则判定为穿越报文。

3 验证

为了验证检测程序的正确性，本实验在江苏省网边界采集了 3 组全报文，利用全报文匹配的方式精确判定每个报文是否为 DNS 报文，以此来做标准答案。具体的验证验证流程如下：

（1） 在江苏省网边界抓取时长为十几分钟的全报文；

（2） 调用 libbind 库中的 ns_initparse 函数进行判定,该函数能够从报文的结构上判断出报文是否为 DNS 报文，准确性高；

（3） 判断答案存储在一个字节中，0 表示否，1 表示是，将所对应的报文体截取 60 个字节，组成一个 61 字节的整体；

（4） 将所有全报文组成一个以 61 字节为单位的二进制文件；

（5） 将作者设计的 DNS 报文检测程序应用在第（4）步中得到的文件；

（6） 将答案与标准答案相比较。

验证的结果如表 3.1 所示，查准率高达 99.99%且无误判数，证明第二节提出的 DNS 检测方法精准率高，具有实际意义，能够将其作为研究方法。

表 3.1 验证结果

组号	53 端口报文数	DNS 报文数	查准数	误判数	查准率	误报率
1	1352369	370889	370871	0	99.99%	0
2	13150211	6110352	6110036	0	99.99%	0
3	12643756	6064377	6064126	0	99.99%	0

4 UDP53 端口流量分析

为了能够分析出 53 端口流量中除了 DNS 服务以外的主要用途，本文利用 DNS 报文检测程序设计了一个 UDP53 端口流量分析系统。系统的主要流程如图 4 所示，其中 result.txt 记录的是报文个数，DNS 报文个数以及 UDP53 报文个数等一些计数量，为了到时算出它们之间的各种比例。correct.txt 中记录的真正的 DNS 服务器列表，而 error.txt 中记录的就是问题 IP 的列表以及它们使用 53 端口的次数。

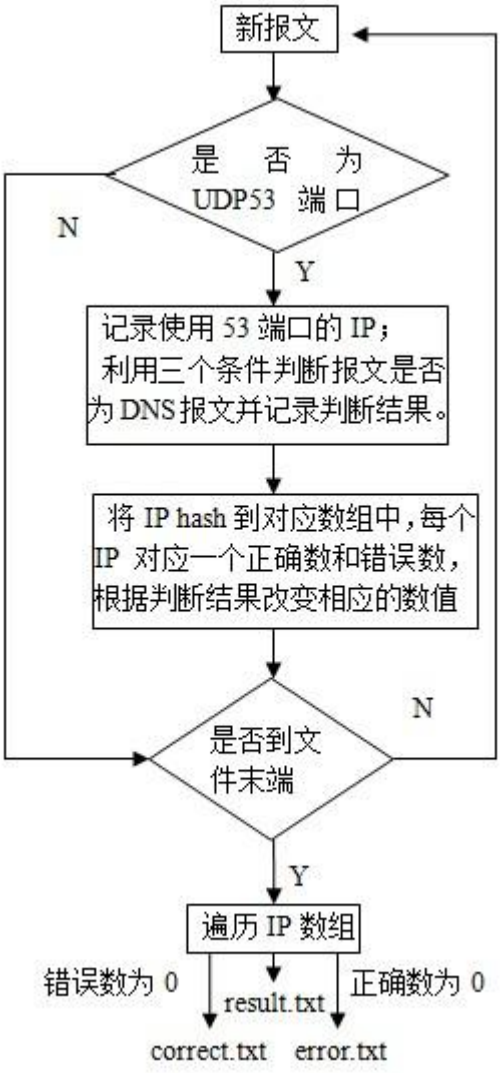

图 4 系统流程图

5 实验结果及分析

5.1 数据集

本文的测试数据集为从 2005 年到 2010 年，来自江苏省网边界采集器的 IP Trace。具体测量数据见表 5.1，表 5.1 列出了数据集采集线路，采集时间，时长以及相应包的数量。其中采集线路是由于采集器在采集时会根据原宿 IP 地址将报文分流成四组，采集线路表示的就是这四条分流。

从表 5.1 可以看出，UDP 报文占总报文的比例从 2005 年的 6.99%上升到 2008 年的 50%，这说明近年来出现的多种基于 UDP 协议的网络应用(例如：P2P、PPStreaming 和 BitTorrent 等)已经开始被广泛使用，由于这些协议大部分与视频流以及大文件传输流有关，所以 UDP 的比例才会有这种很大幅度的提升。根据这些协议本身来说，它们一般使用的是大端口而不会使用一些公认的端口（例如 53）。2005-2008 的数据中 DNS/UDP53 的数值随着 UDP/IP 的增大而变小也正是体现了这一点。因为使用非 53 端口的 UDP 报文增多，而使用 53 端口的 DNS 报文却增幅很小，所以导致比例越来越小。但是 2009 年和 2010 年的 UDP53/UDP 的数值相对于前几年有很大幅度的增长，而根据网络的实际情况，DNS 报文是不会有这么大的涨幅，说明这些“DNS 报文”中可能有大部分报文是使用 53 端口的伪 DNS 报文。本文提出的检测方法就能还原这些报文的真实情况。

表 5.1 数据集

采集线路	采集时间	时长	IP 包/10^6个	UDP 包/10^6个	UDP53/个	UDP/IP	UDP53/UDP
全部	2005-11-10	14:00-15:00	2342	164	5929217	6.99%	3.606%
全部	2006-12-31	14:00-15:00	1662	373	5315218	22.44%	1.423%
1	2007-10-30	14:00-14:40	170	38	510556	22.53%	1.332%
1	2008-12-20	14:00-15:00	724	340	1993835	46.96%	0.585%
1	2009-12-18	14:00-16:00	1896	923	132416594	48.68%	3.553%
1	2010-07-18	14:00-16:00	1141	435	31277250	38.12%	7.178%

5.2 结果分析

表 5.2 列出了测试结果，分别列出了 53 端口 UDP 报文的个数、通过检测判断为 DNS 报文的个数以及两者的比例。DNS/UDP53 的数值在 2005-2006 年都处于 99.5%以上，2007-2008 年有小幅下降，2009-2010 年跌落到了 3.553%和 11.087%，说明从 2009 年起有大量的非 DNS 报文在使用 53 端口。

表 5.2 测试结果

采集线路	采集时间	UDP53/个	DNS/个	DNS/UDP53
全部	2005-11-10	5929217	5900105	99.509%
全部	2006-12-31	5315218	5312967	99.970%
1	2007-10-30	510556	499081	97.752%
1	2008-12-20	1993835	1933694	96.983%
1	2009-12-18	132416594	4705893	3.553%
1	2010-07-18	31277250	3468021	11.087%

以 2010-17-18 的 IP Trace 为例，本文简单分析了这些穿越报文。问题 IP 如表 5.3 所示，根据表的形式，读者不难看书这些 IP 被分为三类。一类为一组属于同一个 C 类的 IP 地址，一类为迅雷离线下载服务器（IP 太多，未列出），另外一类为其他剩下的未知用途的 IP。表中数据显示第一类 IP 被使用的频率达到 88%以上，几乎占据了所有的非 DNS 报文。通过验证，得知这几个 C 类地址是某公司构建在江苏省教育科研网上的主机的 IP，这些主机使用 53 端口传输视频流，以此来穿越边界路由器。第二类 IP 是通过验证确定为迅雷离线下载服务器地址的 IP，它们占 UDP53 端口报文的 0.011%，数量比较小。由于受到 68 字节的限制，对这些报文的分析不能继续深入，但是可以肯定的是这些报文无任何攻击行为。第三类 IP 和第二类 IP 一样所占比例非常小，但是无法确定它们的具体用途。

表 5.3 问题 IP 列表

问题 IP	使用次数	总计	占 UDP53 的比例
C 类地址一	6116615		
C 类地址二	5996624	27794579	88.865%
C 类地址三	8160638		
C 类地址四	7520702		
迅雷离线下载服务器	3503	3503	0.011%
未知		11147	0.035%

6 结束语

本文通过查准率高达 99.99%的 DNS 报文检测程序对 JSERNET UDP53 端口的报文进行了流量分析。测量数据分别分布在 2005-2010 年，从测量结果可以看出，使用 53 端口穿越的报文在 2009 年增多，使得 2009 年和 2010 年的数据中 DNS 报文占 UDP53 端口报文的比例下降到 3.553%和 7.178%。为了了解是哪些 IP 在使用 53 端口，它们是否有一些特性，本文还对那些非 DNS 报文进行了研究，找到了其中的问题 IP 并了解了其用途，为以后对端口的屏蔽工作提供了依据。

参考文献

[1] 田 杰, 谷大武, 陆海宁.预防缓存中毒的 DNS 报文校验方案[J].通信技术,2010,08(43):146

[2] 闫伯信,方淇兴,李斌,王盎.DNS 欺骗攻击的检测和防范[J].计算机工程,2006.11,21(32):130

[3] W. Ricard Steven 著;范建华,胥光辉,张涛等译. TCP/IP 详解卷 1：协议[M].北京:机械工业出版社,2000.4P142-145

[4] IP TASCM 使用手册.2009.