

TCP 与 UDP 网络流量对比分析研究*

张艺濒^{1,2a,2b}, 张志斌^{1,2b}, 赵咏^{1,2a,2b}, 郭莉^{1,2b}

(1. 中国科学院 计算技术研究所, 北京 100190; 2. 中国科学院 a. 研究生院; b. 信息内容安全技术国家工程实验室, 北京 100049)

摘要: 网络带宽不断增长,越来越多的音/视频、在线游戏等应用成为网络空间的主体。基于实时性考虑,这些新兴应用协议多选择 UDP 作为其底层的传输协议,使得 UDP 流量呈上升趋势,而以往的流量测量工作一般基于 TCP 进行,忽略了 UDP 协议。对国内某骨干网流量进行了连续 12 h 的在线测量,在传输层和应用层分别对 TCP 和 UDP 及其应用层协议的流的总数、长度分布、持续时间分布、流的速度分布等进行了详尽的分析,并对 TCP 和 UDP 的应用层协议流的大小、长短、快慢作了详细的分类。为网络流的分类技术、网络行为发现、网络设计等提供了数据支持。

关键词: TCP; UDP; 协议特征; 流量分类; 端口

中图分类号: TP393.09 文献标志码: A 文章编号: 1001-3695(2010)06-2192-06

doi:10.3969/j.issn.1001-3695.2010.06.056

Comparative analysis on TCP and UDP network traffic

ZHANG Yi-bin^{1,2a,2b}, ZHANG Zhi-bin^{1,2b}, ZHAO Yong^{1,2a,2b}, GUO Li^{1,2b}

(1. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China; 2. a. Graduate School, b. National Engineering Laboratory for Information Security Technologies, Chinese Academy of Sciences, Beijing 100049, China)

Abstract: With the increase of network bandwidth, more and more new applications like audio, video and online games become the major force in network traffic. Based on real-time considerations, these new applications mostly used UDP as transport layer protocol, which made UDP traffic increase. While previous traffic measurements were generally based on TCP and ignore UDP protocol. This paper made a continuous 12-hour on-line measurement on a backbone environment. Firstly, it did a detailed analysis on TCP and UDP in transport and application layer, including total number of the flow, length distribution, duration distribution, and flow rate distribution. Secondly, gave a detailed classification on application layer protocols according to flow size, length, and speed. Our analysis and conclusions can present a data support for other studies as the traffic classification, network behavior analysis, and the network design etc.

Key words: TCP; UDP; protocol characteristics; traffic classification; port

0 引言

随着网络带宽不断增长,网络行为模式日益复杂,各种新网络应用的产生,使人们对于网络的依赖性越来越强,同时也对网络提出了更高的要求。传统的基于文字和图片的网络服务已经不能满足人们的需求,越来越多的音/视频、在线游戏等内容生动且互动性强的网络应用开始成为网络空间的主体。根据 2008 年 IResearch 咨询公司通过 Cisco System 提供的数据统计结果,P2P 和在线视频服务流量已经超过了网页浏览、电子邮件这些传统的网络服务,其正在占用越来越多的网络带宽。以 2008 年为例,在线视频(包括 PC 和电视终端)和 P2P 流媒体产生的数据流量将占据全球互联网总流量的 75.9%,达到 4 034 PB,而预计到 2012 年这一数字会上升至 81.2%。显然,以网络视频为主的新兴网络应用服务将成为未来带宽增加的最大推动者和受益方。

一般来说,网络应用服务在选择传输层协议时有两大选择,即 TCP 或 UDP。TCP 是一种面向连接的、具备可靠传输特性的协议;UDP 是一种面向无连接的传输协议。曾占据网络中绝大部分流量的 TCP 目前仍然是网络流量的较大组成部分,但 UDP 流量伴随着新型网络服务的发展而迅猛增加。由于音/视频通信、在线游戏等服务数据冗余性较强,对个别数据包的丢失不敏感,但是都具有较强的实时性要求,而且希望占用尽量少的管理资源从而可以支持更多的同时在线用户,目前这些新兴的网络服务越来越多地选择 UDP 作为其底层的传输协议。

传统的测量工作都是基于 TCP 进行的,忽略了当前网络流量中越来越重要的 UDP,因此测量结果并不能完全代表网络实质。为此,本文将对高速网络中 TCP 和 UDP 流进行测量研究,分析对比两者的统计特性,促进对当前网络流量中 TCP 和 UDP 流的理解,为网络和应用服务的设计提供借鉴。本文

收稿日期: 2009-11-09; 修回日期: 2009-12-23 基金项目: 国家“973”计划资助项目(2007CB311100)

作者简介: 张艺濒(1985-),女,江西景德镇人,硕士研究生,主要研究方向为信息安全、流量分类、UDP 测量等(zhangyb85@gmail.com); 张志斌(1978-),男,山东济南人,助理研究员,主要研究方向为网络流处理、网络测量等; 赵咏(1983-),男,河北石家庄人,博士研究生,主要研究方向为信息安全、流量分类、P2P 测量; 郭莉(1969-),女,北京人,研究员级高级工程师,博士研究生,主要研究方向为算法设计与分析、数据流管理、网络与信息安全。

有下述两点主要贡献:a)在传输层和应用层分别对TCP和UDP及其应用层协议的流的总数、长度分布、持续时间分布、流的速度分布等作了详尽的分析;b)对TCP和UDP的应用层协议流的大小、长短、快慢作了详细的分类。

1 相关工作

面向网络层面的流量测量工作在20世纪80年代就已经起步,较早的研究工作多是以数据包(packet)作为理解网络流量行为的基本单位(building block)。Leland等人^[1]通过对1989年8月~1992年2月间数以亿计的以太网数据包的统计分析,发现以太网流量具有自相似(self-similar)特性。但是正如Clark^[2]指出的,基于包级别流量测量由于粒度小,在很多方面无法满足对于网络流量理解的需要。几乎在相同时期,Jain等人^[3]也提出了相似的看法,他们提出了一种包列模型(packet train)。后续的很多研究^[4,5]将包列模型扩展到传输层或者应用层分析中,或者专门对TCP流进行分析^[6,7]。Claffy等人^[8]在包列模型的基础上根据流量在时间和空间上的局部性特征定义了带参量的流(flow)模型。

从2000年开始,面向流级别的流量测量研究逐渐成为网络测量领域的一个热点问题。Fang等人^[9]通过对骨干网流量数据的研究发现,在AS之间占总数9%的流承载了90%的流量,这也就是在网络流领域所谓的“老鼠与大象”(mice and elephant)现象。Brownlee等人^[10]对网络流的持续时间分布进行了测量研究,发现占总量45%的网络流在持续时间上小于2s,而不到2%的网络流在持续时间上超过15min并且承载了50%以上的流量,即所谓的“蜻蜓与乌龟”(dragonflies and tortoise)现象。Sarvotham等人^[11]通过测量指出流量的突发一般源于少数几个承载流量大的网络流。另外Lan等人^[12]还根据流速的情况将网络流划分为“猎豹与蜗牛”(cheetah and snail),并研究了流的各种属性之间的相互关系。

自2000年以来,以具体应用为背景的网络测量研究主要集中在音/视频、即时聊天、P2P、在线游戏等新的网络服务上,这些测量结果显示UDP正在逐渐成为新兴网络服务进行数据传输的主要选择,其重要性正在逐渐赶上或者超过TCP。Mena等人^[13]发现60%~80%的音频数据流都是通过UDP进行传输,而只使用TCP进行控制命令的传输。Sripinidkulchai等人^[14]对某著名的内容分发网络进行统计,其结果显示UDP在音/视频领域占据绝对的优势。另外,近几年对于以Skype为代表的流量研究^[15]发现,目前的VoIP服务也主要采用了UDP作为传输协议。Dewes等人^[16]指出以IRC和Web为基础的聊天系统一般采用TCP,而其他类型的聊天系统如ICQ和AIM则普遍使用UDP。Feng等人^[17]指出大量第一人视角游戏多用UDP实现。在发展迅速的P2P设计方面,UDP通信也占有相当的比例。Karagiannis等人指出目前主流P2P协议一般都可以使用TCP或者UDP进行数据传输^[1]。

然而基于流的网络测量也有着如下的问题:目前的测量方法多是基于TCP设计的,网络设备也是TCP友好型。随着新兴网络应用的兴起,越来越多的服务开始考虑将数据传输的工作交给UDP完成,因此在流量的成分比例上TCP也不再像以前一样占据绝对优势。网络流量测量的主要目的是要增进人

们对于流量特征的认识,而流量特征是随着所承载应用的变化而变化的,因此,有必要对UDP流的特征进行深入研究,从而指导网络设计向着更符合应用需要的方向发展。为此,本文作了针对TCP和UDP的协议对比分析,以期找出两种传输协议之间的共性和差异,更好地指导网络设计和网络维护。

2 实验环境

本实验为实时测试,数据来源于国内某骨干路由器,带宽为1000Mbps,测试时长为12h,从正午12点至凌晨12点。该测试时间覆盖了网络流量较大的繁忙时段和网络流量较小的空闲时段。

本实验分为如图1所示的五个模块,分别是流量采集、流量分类、流表管理、超时策略和流信息统计。流量采集模块在线采集网络数据包,对数据包进行分层分析、IP分片重组等,再经过流量分类模块,运用特征识别等方法分析出当前数据包属于哪种传输层和网络层的协议,进入流表管理模块;流表管理运用超时策略维护一个TCP活动流表和一个UDP活动流表,当发现流结束或者超时,则统计计算该流的各项参数信息。

本文在包级别和流级别均作了测量。目前广泛采用的定义流的方式是五元组^[4],即拥有相同源IP地址、目的IP地址、协议号、源端口号和目的端口号的一系列数据包的集合。Claffy等人^[8]为实现新建连接数与活动连接数均较小,达到资源消耗总量较小的目的,提出主干网数据流超时时间以64s为宜,此后网络设备提供商均基于此使用64s作为默认超时值。本文借鉴了上述方法,提出本实验中TCP和UDP流的定义。

定义1 拥有相同的四元组(源地址、目的地址、源端口、目的端口)的一系列在一定超时时间内(64s)连续到达的TCP数据包,称为一个TCP流。

定义2 在两个终端之间且拥有相同端口号的一系列在一定超时时间内(64s)连续到达的UDP数据包,称为一个UDP流。

3 基于传输层的测量

从测量对象角度看,如果把流作为理解网络流量行为的基本单位,那么与网络资源占用相关的流的若干基本属性特征也是理解网络流量的重要指标,这其中包括流的总数、长度分布、持续时间分布、流的速度分布等。对这些指标的测量增进了人们对于网络流量的理解和认识。流的这些基本属性特征的测量结果往往能够为网络的设计提供重要的参考,如在路由器设计中,关于缓冲大小设计、包转发策略、流调度算法等方面都参考了网络流测量方面的一些研究成果。

3.1 对包数、字节数、流数的统计

在对数据分析后,可以得出TCP与UDP包数在12h内的变化情况(图2),同理得出TCP和UDP字节数和流数的走势图(图3、4)。

由图2~4综合分析可以看出:TCP在包数上明显比UDP要多;UDP的字节数比TCP要多;TCP流数与UDP总体相当,但随着时间和流量的变化,流数不稳定。从图2和3中可以发现,在正午12~18点,UDP的包数明显小于TCP,但UDP的字节数却比TCP高,这说明,在这个时段内,UDP的包所含字节

数偏大,由此可以推断,网络繁忙时,主要由 UDP 流量占用带宽。在 CAIDA2009 年的一篇报告^[18]中,在 2003 年,UDP 无论在包数、字节数还是流数上,只占网络流量的 20% 以下。从 2006 年开始,UDP 流量在逐年增加,到 2009 年初,UDP 在流级上已经占有了巨大优势,是 TCP 的 2.6~3 倍。

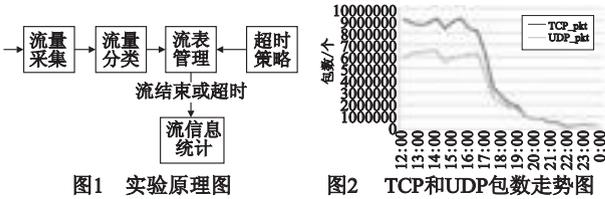


图1 实验原理图

图2 TCP和UDP包数走势图

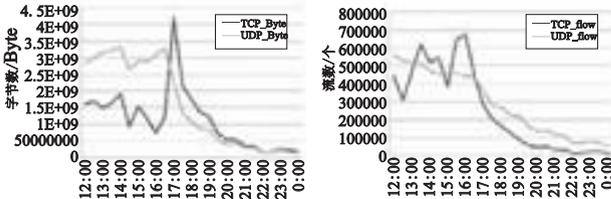


图3 TCP和UDP字节数走势图

图4 TCP和UDP流数走势图

随着新兴网络协议的日益普及,网络流量的组成发生了很大的变化,几年前占用网络带宽 10% 左右的 UDP,如今在网络流量中的比重越来越大,从各个方面看都占据了一半的网络资源,成为了不可忽视的一部分。

3.2 对流的平均包数、字节数、速率以及流持续时间的统计

TCP 和 UDP 的字节数、包数是数据包级别的测量,以数据包为基本粒度对流量进行研究丢弃了很多宏观上的重要信息,因此有必要在流级别上进行更详细的测量。本文选取了流的平均包数、平均字节数、平均速率和平均持续时间这四个参数进行统计。

首先给出文章中使用的字符及符号含义,如表 1 所示。

表 1 符号表

字符	含义
f	流数
p_i	流 i 的包数
b_i	流 i 的字节数
r_i	流 i 的速率
s_i	流 i 的第一个包的到达时间
e_i	流 i 的最后一个包的到达时间

3.2.1 流的平均包数

流的平均包数的定义如下:

$$avg_pkt_perflow = \frac{\sum_{i=1}^f p_i}{f}$$

经过实时测量,得到 TCP 和 UDP 的流的平均包数,如图 5 所示。TCP 流的平均包数在不同的网络情况下数目不同,从 12:00~17:00 流的平均包数在 20~35,并且随着网络流量的减少,在 18:00 之后,流的平均包数稳定在平均每个流 35 个包。UDP 流的平均包数在不同的网络情况下数目比较稳定,流的平均包数在 15 个左右,而 TCP 流的平均包数比 UDP 要多,说明从总体上看 TCP 流偏向于大流,而 UDP 则偏向于小流。

3.2.2 流的平均字节数

流的平均字节数定义如下:

$$avg_Byte_perflow = \frac{\sum_{i=1}^f b_i}{f}$$

经过实时测量,得到 TCP 和 UDP 的流的平均字节数,如图

6 所示。

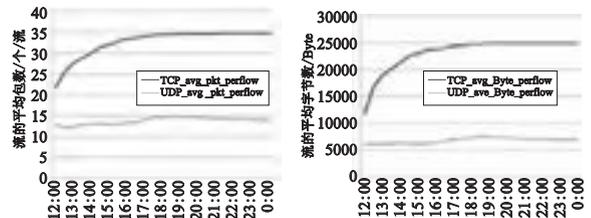


图5 TCP和UDP的流的平均包数 图6 TCP和UDP的流的平均字节数

TCP 流的平均字节数在不同的网络情况下不同,从 12:00~17:00 流的平均字节数在 10 000~25 000 Byte,随着网络流量的减少,在 18:00 之后,流的平均字节数稳定在平均每个流 25 000 Byte。UDP 流的平均字节数在不同的网络情况下比较稳定,流的平均字节数在 5 000 Byte 左右。TCP 流的平均字节数比 UDP 流多,结合图 5,每一个 TCP 流包含的包数多,且字节数也多。TCP 流的平均包数是 UDP 的 2.5 倍左右,但 TCP 流的平均字节数却是 UDP 的 4.5 倍左右。这表明,TCP 中每个包所占的字节数也比 UDP 多,说明 TCP 流较 UDP 而言是大流。

3.2.3 流的平均速率

流的速率指的是流的总字节数与流的持续时间的比值,这一比值可以反映流的快慢,定义如下:

$$avg_rate_perflow = \frac{\sum_{i=1}^f (b_i / (e_i - s_i))}{f}$$

经过测量,可以得到 TCP 和 UDP 的流的速率,如图 7 所示。TCP 流的平均速率在不同的网络状况下比较稳定,维持在 250 Bps,表明在各种情况下,TCP 的使用是比较均衡的。UDP 流在网络流量较大的 12:00~18:00,流的平均速率稳定在 150 Bps,但在 18:00 之后,总的网络流量减小,而 UDP 流的平均流速却有所提高。这表明在夜间高速的 UDP 流的成分有所提升,如在夜晚观看 P2P 在线视频的人数增加,导致 UDP 的平均流速增加。对比而言,TCP 流的速率明显比 UDP 快,作为大流的 TCP,同时也是快流。

3.2.4 流的平均持续时间

流的持续时间指的是流的第一个包到达的时间和最后一个包到达的时间间隔,定义如下:

$$avg_duration_perflow = \frac{\sum_{i=1}^f (e_i - s_i)}{f}$$

经过测量,可以得到 TCP 和 UDP 流的平均持续时间如图 8 所示。

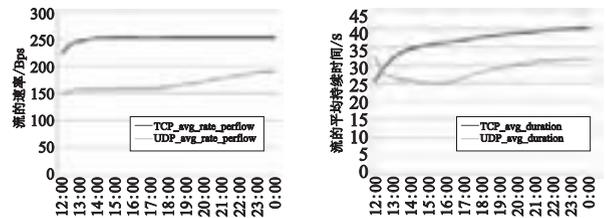


图7 TCP和UDP的流的平均速率 图8 TCP和UDP的流的平均持续时间

TCP 流的平均持续时间随着网络流量的减小而缓慢增加,一定程度上可以反映出在白天人们对信息的索取是简单快速的,而到了夜间,人们的生活节奏变缓,更加倾向于慢慢地浏览信息。UDP 流在网络流量较大的 12:00~18:00,流的持续时间不稳定,但在 18:00 之后,总的网络流量减小,UDP 流的平

均持续时间却有所提高。这表明在夜间长流的 UDP 流的成分有所提升,如在夜晚观看 P2P 在线视频的人数比例增加,导致 UDP 的平均持续时间增加。对比而言,TCP 流的平均持续时间大部分情况下都比 UDP 要长。

结合这四个参数总体上看,TCP 是大流、快流和长流,对比网络流量较大的日间和网络流量较小的夜间,TCP 在夜间呈现更大更长的特征。UDP 相比 TCP 是小流、慢流和短流,对比网络流量较大的日间和网络流量较小的夜间,UDP 在夜间呈现更快更长的特征。

3.3 对端口的统计

对 12 个小时的端口号进行统计,分别得到 TCP 与 UDP 含量较多的前五个端口号,如表 2 所示。

表 2 TCP 和 UDP 前五个端口号

协议	商品				
	1	2	3	4	5
TCP	80	25	443	4 662	1 301
UDP	8 000	15 000	53	7 600	80

TCP 中 80 端口为 HTTP,表明 HTTP 仍占据着 TCP 的主要部分。占据 TCP 端口号第二位的 25 端口为 SMTP。443 端口为网页浏览端口,主要用于 HTTPS 服务;4662 端口为 Emule (电驴)协议的标准端口号。

对 UDP,8000、15000 和 53 端口分别对应着 QQ、迅雷和 DNS 协议。

4 基于应用层的测量

在对 TCP 和 UDP 作了初步的测量之后,可以得出一些关于 TCP 和 UDP 的比较结果,然而,这些结果是针对所有协议而言的。在应用层协议日益丰富的现在,应用层协议的统计特征千差万别。为了更好地体现 TCP 与 UDP 之间的差别,在此使用了端口识别和特征提取的方法,识别了 10 余种网络中含量较多的应用层协议,并对其进行了详尽的分析。

针对协议的测量工作分为两个部分,即协议识别和应用层协议特征统计。

4.1 协议识别

4.1.1 TCP 的应用层协议识别

由于使用 TCP 传输协议的应用层协议通常都有固定的端口,在此本文使用端口识别的方法分析 TCP 流,如表 3 所示。

表 3 TCP 应用层协议及其相应的端口号

端口	协议					
	HTTP	HTTPS	FTP	TELNET	SMTP	POP3
Port	80	443	21/20	23	25	110

进一步对各种 TCP 的端口号进行分析,80 端口号的连接(HTTP)数目在流量中占主要部分,同时 25 端口号的连接(SMTP)数目也较多。

4.1.2 UDP 的应用层协议识别

目前 UDP 流承载了丰富的应用类型,虽然这些应用类型都具有一定的实时性要求,但是不同的应用类型之间又会有比较大的差异。例如音/视频的实时性要求就比较高,但是对于数据包丢失基本可以容忍;而对于基于文本的聊天系统来说,数据包丢失可能会造成信息的不完整,但是却可以容忍比较高的信息传输延迟。理解 UDP 流中的应用成分将有助于按照应

用需求进行流量工程,并根据网络资源情况提供服务质量(QoS)保证。

L7-filter 和 Open DPI 等开源软件提供了众多协议的特征信息。本文借鉴了部分协议特征及反汇编知识分析了网络中含量较多的 10 余种应用层协议,如表 4 所示。

表 4 UDP 应用层协议分类表

协议名称	占总包数比例/%	占总流量比例/%	占总流数比例/%
UNKNOWN	51.2	58.9	26.9
XUNLEI	12.4	15	1.03
PPSTREAM	6.78	8.7	0.02
BITTORRENT	13.8	7	46.5
PPLIVE	3.0	6.2	0.02
DNS	8.6	2.4	15.8
MESSENGERSERVICE	1.1	1.7	6.1
QQLIVE	1.6	0.4	0.36
SOPCAST	0.2	0.3	0.01
EDONKEY	0.9	0.3	2.7
QQ	0.1	0.04	0.05
NTP	0.2	0.04	0.4

4.2 各应用层协议对流的平均包数、字节数、速率以及流的持续时间的统计

4.2.1 流的平均包数

经分析得到 TCP 和 UDP 的各种应用层协议的流的平均包数,如图 9、10 所示。



图 9 TCP 各应用层协议流的平均包数

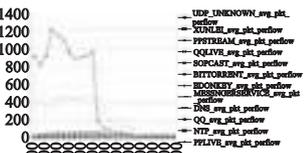


图 10 UDP 各应用层协议流的平均包数

由图 9 可以看出,HTTP 流的包数较多,TELNET 流的平均包数较小,且各种应用层协议的包数差异较大。注意到无论在流量较多的日间还是流量较少的夜间,SMTP、FTP 和 HTTPS 协议流的平均包数比较稳定。由图 10 可以看出,PPLIVE 和 PPSTREAM 是典型的大流,日间流的平均包数在 400 以上,当用户量减少的夜间,流的平均包数才减小至 100 个包左右,这是视频流的典型特征。

4.2.2 流的平均字节数

经分析得到 TCP 和 UDP 的各种应用层协议的流的平均字节数,如图 11、12 所示。

HTTP 流的平均字节数远高于其他协议,表明 HTTP 流是典型的大流。SMTP 和 FTP 的平均字节数在任何情况下都较为稳定,表明这两种协议有着稳定的特性。



图 11 TCP 各应用层协议流的平均字节数

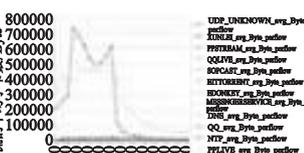


图 12 UDP 各应用层协议流的平均字节数

图 12 与图 10 类似,PPLIVE 和 PPSTREAM 是典型的大流,日间的流的平均字节数远远高于其他协议,当用户量减少的夜间,流的平均字节数才有所减小,这是视频流的典型特征。

4.2.3 流的平均速率

经分析得到 TCP 和 UDP 的各种应用层协议的流的平均速率,如图 13、14 所示。

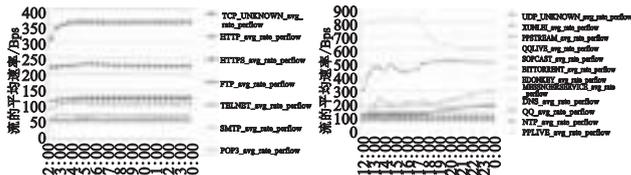


图13 TCP各应用层协议流的平均速率

图14 UDP各应用层协议流的平均速率

由图 13 可知,无论在何种网络(流量大小)条件下,TCP 的各种应用层协议的流的平均速率都是很稳定的。HTTP 的平均速率为 360 Bps,HTTPS 为 225 Bps,POP3 为 115 Bps,SMTP 为 110 Bps,TELNET 为 75 Bps,FTP 为 60 Bps,其他未分类的协议的平均流速率为 120 Bps。

利用这一特征,可以得到区分 TCP 各种应用层协议的新方法。

PPLIVE、PPSTREAM、SOPCAST、XUNLEI 等协议在网络流量不稳定的情况下,速率也不稳定。然而,其他协议的速率均较为稳定。BITTORRENT 协议的平均速率为 180 Bps,DNS 为 130 Bps,QLLIVE 为 120 Bps,QQ 为 100 Bps,EDONKEY 为 90 Bps,NTP 为 90 Bps。这一发现可以为应用层协议分类带来新的思路:利用协议的速率特征(或者多种特征的集合)来识别不同的应用层协议。

4.2.4 流的平均持续时间

经分析得到 TCP 和 UDP 各种应用层协议的流的平均持续时间,如图 15、16 所示。



图15 TCP各应用层协议流的平均持续时间

图16 UDP各应用层协议流的平均持续时间

虽然 HTTP 每个流包含的包数和字节数都最高,但是 HTTP 流的持续时间并不是最长的,也就是说 HTTP 流倾向于大流,但非最长流。POP3 协议每个流包含的包数和字节数都较高,但其持续时间最短,且比较稳定,维持在 8 s 左右,表明 POP3 协议是典型的短流。

PPSTREAM 协议不仅是大流,也是长流;QQ 协议的流的包数和字节数都不多,速率也不快,然而 QQ 流的持续时间较长,为长流;DNS、NTP、SOPCAST 等都是短流。

4.3 结论

为进一步分析各种应用层协议流的属性特征,本文借鉴现有文献中对流的大小、快慢区分的标准,为各应用层协议流的特征分类。根据文献[10],流的持续时间小于 2 s 的为超短流,在 2 s ~ 15 min 的为短流,15 min 以上的为长流。根据文献[19]中的定义,流的包数大于 6 的为大流,包数小于 6 的为小流;将平均速率高于 200 Bps 的流定义为快流,将平均速率低于 100 的流定义为慢流。

根据流的大小、快慢和长短的定义,对各种 TCP 和 UDP 应用层协议的分析如表 5、6 所示。

表 5 TCP 应用层协议分析

TCP	流的大小	流的快慢	流的长短
HTTP	大	快	长
HTTPS	大	快	长
FTP	中	慢	长
TELNET	小	慢	长
SMTP	中	中	长
POP3	大	中	短

表 6 UDP 应用层协议分析

UDP	流的大小	流的快慢	流的长短
XUNLEI	大	中	长
PPSTREAM	大	快	长
QLLIVE	大	慢	中
SOPCAST	大	快	中
BITTORRENT	小	中	短
EDONKEY	小	慢	短
MESSENGERSERVICE	大	快	中
DNS	小	中	短
QQ	大	慢	长
NTP	小	慢	短
PPLIVE	大	快	中

由表 5 和 6 可以看到,不同应用层协议流有不同的属性特性。分析应用层协议所属类别后,可以发现如下特征:

a) 网页类协议(如 HTTP、HTTPS)的流是大流、快流和长流。这可以显示用户使用浏览器的部分浏览习惯,大多数用户倾向于长时间浏览大量网页信息。

b) 视频类协议(如 PPSTREAM、PPLIVE、SOPCAST 等)的流是大流、快流和中长流。为了更清晰地观赏视频信息,视频类应用层协议传输网络数据,速度较快、字节数较多。若用户对当前的节目满意时,不会频繁地更换频道,导致视频类协议的流较长。

c) 聊天类协议(如 QQ)的流是大流、慢流和长流。当用户登录聊天软件客户端后,数据传输的快慢大部分情况下取决于用户聊天速度的快慢。聊天类协议的流的特性就是流较长较大,但速度很慢。

5 结束语

流量测量问题作为信息安全领域一个重要的研究方向已经得到越来越多人的关注,然而目前的网络测量大多针对 TCP。本文在传输层上对 TCP、UDP 作了详细的测量和统计工作。同时在应用层协议分类技术的基础上,对网络中出现较多的应用层协议作出统计分析,并分析总结了当前网络中比较热门的服务类型和广泛使用的应用协议的流量特征,得出的结论可以应用于测量统计和分析。网络流的分类识别都是如今网络流研究的重点,然而要完全识别出网络中存在的所有应用层协议是困难的。利用本文的分析方法可以初步定位当前网络流属于哪种服务类型,对网络流量检测与控制有着重大的意义。后续将会利用本文得到的流的各种特征,利用统计学习的方法,对其进行更详尽的分析,以期得到协议识别的新方法。

参考文献:

[1] LELAND W E, TAQUU M S, WILLINGER W, et al. On the self-similar nature of Ethernet traffic[J]. IEEE/ACM Trans on Networking, 1994, 2(1) : 1-15.

- [2] CLARK D D. The design philosophy of the DARPA Internet protocols [J]. *ACM SIGCOMM Computer Communication Review*, 1995, 25(1):102-111.
- [3] JAIN R, ROUTHIER S A. Packet trains-measurements and a new model for computer network traffic [J]. *IEEE Journal on Selected Area in Communications*, 1986, 4(6): 986-995.
- [4] ACHARYA M, NEWMAN-WOLFE U, LATCHMAN H, *et al.* Real-time hierarchical traffic characterization of a campus area network [C]//Proc of the 6th International Conference on Modeling Techniques and Tools, for Computer Performance Evaluation. 1992.
- [5] ACHARYA M, BHALLA B. A flow model for computer network traffic using real-time measurements [C]//Proc of the 2nd International Conference on Telecommunications Systems Modeling and Analysis. 1994.
- [6] CACERES R, DANZIG P B, JAMIN S, *et al.* Characteristics of wide-area TCP/IP conversations [J]. *ACM SIGCOMM Computer Communication Review*, 1991, 21(4):101-112.
- [7] ESTRIN D, MITZEL D J. An assessment of state and lookup overhead in routers [C]//Proc of the 11th Annual Joint Conference of the IEEE Computer and Communications Societies on One Worlds Through Communications. 1992:2332-2342.
- [8] CLAFFY K C, BRAUN H W, POLYZOS G C. A parameterizable methodology for Internet traffic flow profiling [J]. *IEEE Journal on Selected Area in Communications*, 1995, 13(8):1481-1494.
- [9] FANG Wen-jia, PETERSON L. Inter-AS traffic patterns and their implications [C]//Proc of IEEE Global Telecommunications Conference. 1999:1859-1868.
- [10] BROWNLEE N, CLAFFY K C. Understanding Internet traffic streams: dragonflies and tortoises [J]. *IEEE Communications Magazine*, 2002, 40(10):110-117.
- [11] SARVOTHAM S, RIEDI R, BARANIUK R. Connection-level analysis and modeling of network traffic [C]//Proc of ACM SIGCOMM Internet Measurement Workshop 2001. San Francisco: [s. n.], 2001.
- [12] LAN K C, HEIDEMANN J. A measurement study of correlations of Internet flow characteristics [J]. *Computer Networks*, 2006, 50(1):46-62.
- [13] MENA A, HEIDEMANN J. An empirical study of real audio traffic [C]//Proc of the 9th IEEE INFOCOM Annual Joint Conference of the IEEE Computer and Communications Societies. 2000:101-110.
- [14] SRIPANIDKULCHAI K, MAGGS B, ZHANG H. An analysis of live streaming workloads on the Internet [C]//Proc the 4th ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2004:41-54.
- [15] BONFIGLIO D, MELLIA M, MEO M, *et al.* Tracking down skype traffic [C]//Proc of IEEE INFOCOM'08. 2008.
- [16] DEWES C, WICHMANN A, FELDMANN A. An analysis of Internet chat systems [C]//Proc of the 3rd ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2003:51-64.
- [17] FENG W, CHANG F, FENG W, *et al.* A traffic characterization of popular on-line games [J]. *IEEE/ACM Trans on Networking*, 2005, 13(3):488-499.
- [18] <http://www.caida.org/research/traffic-analysis/tcpudpratio/> [EB/OL].
- [19] 周明中, 龚俭, 丁伟. 高速网络中基于流速测度的动态超时策略 [J]. *软件学报*, 2006, 17(10):2141-2151.
- [20] GILDER G, TELECOSM G. How infinite bandwidth will revolutionize our world [M]. New York: Free Press, 2000:58-67.
- [21] 中国互联网络信息中心. 第23次中国互联网络发展状况统计报告 [EB/OL]. (2009). <http://www.cnnic.net.cn/uploadfiles/pdf/2009/1/13/92458.pdf>.
- [22] http://www.iresearch.com.cn/html/Consulting/Online_Movie/DetailNews_id_86894.html [EB/OL].
- [23] GUO L, TAN E, CHEN Song-qing, *et al.* Delving into Internet streaming media delivery: a quality and resource utilization perspective [C]//Proc of the 6th ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2006:217-230.
- [24] GUO L, TAN E, CHEN S, *et al.* Analysis of multimedia workloads with implications for Internet streaming [C]//Proc of the 14th International Conference on World Wide Web. New York: ACM Press, 2005:519-528.
- [25] KHALIL K M, LUC K Q, WILSON D V. LAN traffic analysis and workload characterization [C]//Proc of the 15th Conference on Local Computer Networks. 1990:112-122.
- [26] BROWNLEE N, MILLS C, RUTH G. RFC 2722, Traffic flow measurement: architecture [S]. 1999.
- [27] THOMPSON K, MILLER G, WILDER R. Wide-area traffic patterns and characteristics [J]. *IEEE Network*, 1997, 11(6):10-23.
- [28] ROBERTS L G. Beyond Moore's law: Internet growth trends [J]. *IEEE Computer Magazine*, 2000, 33(1):117-119.
- [29] DUFFIELD N, LUND C, THORUP M. Estimating flow distributions from sampled flow statistics [C]//Proc of Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. 2003:325-336.
- [30] ESTAN C, VARGHESE G. New directions in traffic measurement and accounting [J]. *ACM SIGCOMM Computer Communication Review*, 2002, 32(4):323-326.
- [31] MORI T, UCHIDA M, KAWAHARA R, *et al.* Identifying elephant flows through periodically sampled packets [C]//Proc of the 4th ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2004:115-120.
- [32] ESTAN C, KEYS K, MOORE D, *et al.* Building a better netflow [C]//Proc of Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. 2004:245-256.
- [33] DUFFIELD N, GROSSGLAUSER M. Trajectory sampling with unreliable reporting [C]//Proc of IEEE INFOCOM Conference. 2004.
- [34] KUMAR A, XU J, LI L, *et al.* Space-code bloom filter for efficient traffic flow measurement [C]//Proc of IEEE INFOCOM Conference. 2004.
- [35] SCHWELLER R, GUPTA A, PARSONS E. Reversible sketches for efficient and accurate change detection over network data streams [C]//Proc of the 4th ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2004:207-212.
- [36] ESTAN C, VARGHESE G, FISK M. Bitmap algorithms for counting active flows on high speed links [C]//Proc of the 3rd ACM SIGCOMM Conference on Internet Measurement. New York: ACM Press, 2003:153-166.
- [37] ESTAN C, SAVAGE S, VARGHESE G. Automatically inferring patterns of resource consumption in network traffic [C]//Proc of Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications. 2003:137-148.
- [38] ZHANG Hui. Internet video: new wine, old bottle [C]//Proc of Key-note of ICNP'07. 2007.