大规模网络中IP流流速分析*)

吴 桦 周明中 龚 俭

(东南大学计算机科学与工程学院 南京 210096) (江苏省计算机网络技术重点实验室 南京 210018)

摘 要 作为Internet流量模型研究的重要组成部分,IP流流速反映了各种不同应用类型的流量在网络中对实际负载的贡献情况。通过分析和寻找对IP流平均流速产生主要影响的若干关键因子,可以为基于IP流的流量模型提供必要的条件。本文首先分别对IP流的三种主要构成部分:TCP流、UDP流和ICMP流的平均流速从协议分析的角度进行建模,从平均流速模型参数分析中,获得在不同阶段对决定IP流平均流速的若干主要影响因子;然后,使用采集自各种不同时间、不同应用背景和不同负载的大规模高速网络的TRACE作为研究对象,通过试验分析的方法,对所选取的TRACE中不同协议类型的IP流平均流速进行统计分析,检验这些因素在实际网络中在不同流长的情况下对IP流平均流速的贡献,从而验证了本文所提出的IP流平均流速影响因子的可靠性。

关键词 网络行为,IP流流速,网络被动测量,大规模网络,统计学

Analysis of IP Flows Rate in Large-scale Networks

WU Hua ZHOU Ming-Zhong GONG Jian

(School of Computer Science and Engineering, Southeast Univ., Nanjing 210096)

(Jiangsu Province Key Laboratory of Computer Networking Technology, Nanjing 210018)

Abstract As one of most important components of Internet traffic modeling, the IP flow rate can be used to describe the load distribution of different applications in the network. And it is essential condition for IP flow model that some key factors could be found which influenced the flow rate heavily. Firstly, this paper modeling the IP flow rate based on protocol analysis for the three main components: TCP flows, UDP flows and ICMP flows. From the analysis of parameters of those models, the flows rate are found out to be determined by some influence factors. And then, those TRACEs which come from the networks with different time, different areas and different payloads are chosen as study objects. IP flow rate of the different protocol IP flows in these traces are analyzed using statistical method. The influence factors are verified their efficiency for the IP flows rate with different length in the actual networks, and the reliability of the IP flows rate influence factors which are introduced by this paper is also verified.

Keywords Network behavior, IP flows rate, Network passive measurement, Large-scale networks, Statistics

1 引言

随着网络规模的扩大,网络流量建模已经成为网络性能研究最主要的方向之一。通过网络流量建模,可以为理解流量特性,预测网络性能,保障QoS和SLA提供必要的支持。作为网络流量负载的基本度量单位,IP流是指符合特定的IP流规范约束的一系列数据报文的集合,而IP流平均流速是指单位时间内属于特定IP流的报文到达数量或者到达报文的总比特数。研究IP流速模型对路由器中的主动队列管理、服务质量控制,以及网络流量工程等有重要的意义,因此成为网络行为研究的一个重点。

在现有文献中,IP流平均流速的研究主要集中于两个方面:(1)平均流速作为流量的分类[1,3,8];(2)平均流速与其他流特征如流长、流持续时间之间的关系[4~7]。文[1]提出了alpha和beta流量的概念,其中alpha流量指大文件、高带宽传输所产生的流量,而所有不属于alpha的流量均被归纳为beta流量。N. Brownlee, K. Claffy[3]将IP流按照持续时间分为Dragonflies和Tortoises两类,针对他们所测量的TRACE试验分析,发现绝大多数IP流持续时间较短,45%的IP流持续时间小于2s,持续时间小于15min的IP流所占比例大于98%,而剩余不到2%的IP流持续时间加长以至于它们承担了串路50%~60%的比特数的串路负载。文[8]着重介绍一种新颖的超时算法,并使用该算法对IP流流长、流速和流到达等相关特征进行考察,得到流长分布呈重尾、长流流速不稳定,以及流到达存在较多突发等若干结论,但该文只是陈述测量TRACE所得的结果,并没有对形成这些特征的可能原因作深入的分析,而且观测所采用的TRACE数量有限。文[4]通过对TCP流流长和流速的分布特征分析,得出长TCP流的流速和流长存在强相关关系的测量结果,并通过该测量结果来支持一个推断:用户会根据其所处网段的可用带宽状况选择其所传输文件的大小,也就是说用户的选择对长TCP流流量特征起到了决定性作用;文[5,6]同样也研究了流长、流速和流持续时间之间的关系,同时讨论了流内报文到达的突发现象,得出了短TCP流的流速特征主要反映为协议行为的结论。

目前文献研究大多集中于通过TRACE分析和研究,讨论在被研究网络中IP流在平均流速方面存在的现象。尚未有文献从IP流平均流速建模的角度,探讨不同类型的IP流在不同情况下平均流速主要受哪些因素的影响,从而从理论的角度系统地分析IP流平均流速随流长变化的分布状况。

*)本文受国家973计划课题(2003CB314804)、教育部科学技术重点研究项目(105084)和国家863计划(No. 2005AA103011-1)资助。吴 桦 讲师,博士研究生,主要研究方向:网络行为学;周明中 博士,主要研究网络行为学;龚 俭 博导,教授,主要研究网络安全和网络行为。

本文第2部分首先从协议的角度出发，分别对TCP流、UDP流和ICMP流进行流速建模，使用主成分分析方法推测导致IP流流速特征的网络协议行为和用户行为，提出了影响IP流平均流速的若干因子；特别地，由于TCP流传输需要建立可靠连接，所以在连接建立的各个阶段，TCP流平均流速的主要影响因子各不相同；在第3部分，使用来自不同大规模网络不同时段的TRACE作为分析对象，分析不同类型的IP流平均流速随流长的变化状况，从而验证了IP流平均流速主要影响因子的可靠性；第4部分总结了IP流平均流速分布的若干特征；最后给出了关于IP流平均流速影响因子的相关结论，并提出了未来研究的方向。

2 基于协议分析的IP流流速模型

由于IP流主要可以分为TCP流、UDP流和ICMP流，其中TCP流的传输是通过在源宿端点建立串接来保证传输的可靠性，并且通过发现丢包来调整传输的速率，因此从控制系统的角度分析TCP协议属于闭环控制，可以感知网络的使用状况；而UDP流和ICMP流的传输不能感知网络的具体状况，它们平均流速主要与上层应用协议、用户行为相关。

2.1 TCP流传输平均速率模型

由于TCP流在传输时需要建立可靠的串接，因此实际上TCP流的传输可以分为若干个阶段进行。本文将TCP流传输主要分成以下几个阶段作讨论：

a) 串接创建阶段。主要是指交互双方通过三次握手建立串接；

b) 慢启动阶段。在未出现传输报文丢失的情况下，TCP的发送窗口从初始窗口开始随着接收到正常Ack报文的增长而不断增长，发送速率不断增加；

c) 剩余数据阶段。当TCP流中数据传输出现第一次丢包时，则将其以后交互双方之间的数据传输均称为剩余数据传输阶段。

为对TCP流的传输速率进行建模，需要定义若干影响TCP流传输平均传输速率的参数。

RTT：TCP流建立正常串接后往返时延的均值；

RTO：TCP协议规定的初始超时；

W_{SS}：慢启动阶段发送方窗口可达到的最大值；

W_{max}：接收方窗口的最大值；

d：TCP交互双向传输报文数比例（由文[9,10]可知，在TCP每一次交互中，发送方根据发送窗口内的所有报文，接收方在接收到d个报文之后返回一个Ack报文。在具体的TCP不同的实现中，d值是各不相同的。但是目前网络中普遍采用TCP协议实现中d的取值为2，也就是接收方每接收两个报文发送一个应答（delayed ACK）；

p：双向平均丢包率。

2.1.1 串接创建阶段

(1)正常创建串接

根据TCP协议规范[9]，正常创建一个TCP串接平均所需时间为（假设 p_S，p_{SA} 均小于0.5）：

$$E[T_{CE}]=RTT+[(1-p_S)\sum_{i=1}^{\infty}p_S^i\sum_{j=1}^{i}2^{j-1}+(1-p_{SA})\sum_{i=1}^{\infty}p_{SA}^i\sum_{j=1}^{i}2^{j-1}]RTO$$

$$=RTT+\lim_{i\to\infty}[\frac{1+p_S^i(1-2p_S)-(2p_S)^i(1-p_S)}{1-2p_S}+\frac{1+p_{SA}^i(1-2p_{SA})-(2p_{SA})^i(1-p_{SA})}{1-2p_{SA}}]RTO$$

$$=RTT+RTO(\frac{p_S}{1-2p_S}+\frac{p_{SA}}{1-2p_{SA}}) \quad (1)$$

p_S，p_{SA} 分别对应于SYN报文和SYN/ACK报文被丢弃的比例，对用户而言也就是对应于线路上行和下行的丢包率。在边界网络中，由于上下行带宽占用率不一样，p_S，p_{SA} 不相同。但是，由于对于大规模高速网络而言，两者的差别比较小，因此在具体考核时，假设 $P_S=P_{SA}=p$，则式(1)可表达为

$$E[T_{CE}]=RTT+RTO\cdot\frac{2p}{1-2p} \quad (2)$$

由于每次出现超时需要重传一个报文，而正常的TCP建立串接需要双向传输报文数Ninit_CE为2个，则总共传输报文数量可以计算如下：

$$E[N_{CE}]=N_{init\_CE}+\frac{2p}{1-2p}=3+\frac{2p}{1-2p} \quad (3)$$

则此阶段TCP流平均流速为

$$E[V_{CE}]=E[N_{CE}]/E[T_{CE}]=(3+\frac{2p}{1-2p})/(RTT+RTO\cdot\frac{2p}{1-2p}) \quad (4)$$

(2)串接创建失败

由于对方主机或者网络不可达等情况串接请求所传输的SYN报文被连续丢弃，因此网络中存在未能建立正常串接的TCP流。由于每次SYN报文未接收到应答，TCP重传超时增长为原来的两倍，因此这些TCP流传输时间为

$$E[T_{CE}]=\sum_{i=1}^{k}2^{i-1}\cdot RTO \quad (5)$$

其中 k 为源主机尝试发送SYN报文的个数。在具体的实现中，k 值一般为2或者3。而对应创建串接失败的TCP流，发送报文数为 $k+1$ 个：

$$E[N_{CE}]=(k+1) \quad (6)$$

此时TCP流平均流速为

$$E[V_{CE}]=E[N_{CE}]/E[T_{CE}]=(k+1)/(\sum_{i=1}^{k}2^{i-1}\cdot RTO) \quad (7)$$

因此，若未能正常建立串接的TCP流从测量的角度分析属于未应答流，而且持续时间只与RTO相关，传输报文数量非常小。

2.1.2 慢启动阶段

当源宿双方正常建立串接之后，就可以开始传输数据，TCP串接进入慢启动阶段。假设TCP进入慢启动阶段，在发生丢包前，发送方传输 $E[NSS]$ 个报文。如果假设丢包率与发送方行为无关，根据文[10]对TCP交互时间的建模可知，在慢启动阶段实际传输时延为

$$E[T_{SS}]=\begin{cases}RTT\cdot[\log_{(1+1/d)}(\frac{W_{max}}{w_1})+1+\frac{1}{W_{max}}(E[N_{SS}]-\frac{(1+1/d)W_{max}-w_1}{1/d})], E[W_{SS}]>W_{max}\\ RTT\cdot\log_{(1+1/d)}(\frac{E[N_{SS}]\cdot 1/d}{w_1}+1), E[W_{SS}]\leqslant W_{max}\end{cases}$$

其中 w_1 为传输窗口初始大小，$E[W_{SS}]$ 为发送窗口最大值的期望。

由于TCP传输存在delay ACK等传输机制[9~11]，慢启动阶段接受方实际返回的Ack数量为 $\lceil 1/d\rceil\cdot W_i$，而下一次发送窗口将增长为 $W_{i+1}=(1+\lceil 1/d\rceil)W_i$，其中 W_i 为第 i 次发送窗口的大小，W_{i+1} 为下一次发送窗口大小，$\lceil 1/d\rceil$ 为大于 $1/d$ 的最小正整数。由于 d 的取值恒大于1，因此双向实

际传输报文数的期望 $E[N_{SS}]$为

$$(1+1/d)E[N_{SS}]\leqslant E[N_{D\_SS}]\leqslant 2E[N_{SS}] \quad (8)$$

由 TCP 流的双向报文传输数量和传输时间，可以得到在慢启动阶段，TCP 流的平均速率为

$$R(d,w_1,N_{SS},W_{max},RTT)=E[N_{D\_SS}]/E[T_{SS}] \quad (9)$$

因此可知，当 TCP 流处于第一次丢包前的慢启动阶段时，其平均速率是双向传输比例 d、初始发送窗口 w_1、最大接收窗口 W_{max}和平均往返时延 RTT 的函数。

2.1.3 *剩余数据传输阶段*

由 TCP 协议 AIMD 可知，发送窗口随着接收到的正确的确认号增加而增加，直到发生丢包时将其发送窗口减小。文[12]将 TCP 协议类型的数据传输看成是由一系列交替的 Cycle 构成，如图 1 所示。每一个 Cycle 由两个部分组成：慢启动阶段和连续丢包阶段。在慢启动阶段，发送窗口随着接收的正确确认而增长；丢包可能通过两种情况被觉察：重传超时(RTO)和三次重复确认(trIPle-duplicate ACKs)。重传超时情况下，发送窗口将被初始化为 w_1(一般情况下 w_1 值为 1)；在三次重复确认情况下，TCP 发送窗口将减小为原来的 1/2，并启动快速重传。

图 1 中，Cycle1 中丢包是被通过三次重复确认来感知的，因此发送窗口减小为原来的 1/2，并启动快速重传；其他三个 Cycle 都是通过重传超时来感知丢包的，因此发送窗口减小为 1，然后发送窗口进入慢启动阶段。重传超时一般出现在网络拥塞状况比较严重的情况下。

由文[11]对 TCP 流的传输速率模型可知，如果考虑接收窗口的大小，则数据发送方单向传输速率为

$$R'(p,d,RTT,RTO,W_{max})\approx\min\left\{\frac{W_{max}}{RTT},\ \frac{1}{RTT\sqrt{2dp/3}+RTO\min\{1,3\sqrt{3dp/8}\}p(1+32p^2)}\right\}$$

由于双向传输比例为 d，当发送窗口值远大于 d 或者为 d 的整数倍时，TCP 流的平均速率为

$$R(p,d,RTT,RTO,W_{max})\approx\min\left\{\frac{(1+1/d)W_{max}}{RTT},\ \frac{1+1/d}{RTT\sqrt{2dp/3}+RTO\min\{1,3\sqrt{3dp/8}\}p(1+32p^2)}\right\}$$

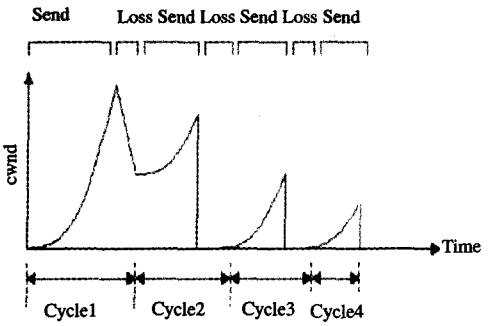

图 1 TCP 拥塞窗口的变化示意图

在实际网络运行中，由于网络传输瓶颈带宽的限制，接收窗口的大小一般不是限制 TCP 流平均速率的主要因素，因此上式可以简化为

$$R(p,d,RTT,RTO)\approx\frac{1}{RTT}\times\frac{1+1/d}{\sqrt{2dp/3}+RTO\min\{1,3\sqrt{3dp/8}\}p(1+32p^2)} \quad (10)$$

因此，TCP 流的平均速率可以看作一个由若干参数共同表示的函数，这些参数包括：平均丢包率 p、双向报传输比例 d、平均传输时延 RTT 和初始重传超时 RTO。

由于在通常情况下，RTT 的值较小[13]。即使是在亚洲/欧洲和美国跨洋串路[14]中，RTT 一般也在 200ms 以下。而 RTO 的取值一般较大，在文[9]中建议 RTO 的取值为 3s(实际观测目前大部分 TCP 的实现也采用这个推荐值)，由此可见 RTO 的值一般要远高于 RTT 的值。如果假设 RTT 和 RTO 为常量，由于公式(10)中 $\min\{1,3\sqrt{3dp/8}\}p(1+32p^2)$ 和 $\sqrt{2dp/3}$ 均为关于 p 的递增函数，但是随着 p 的增加，前者的增长速度远高于后者，因此随着丢包率的增加，RTO 对平均流速影响高于 RTT，即 TCP 流的平均流速也会随着丢包率的增加而迅速下降。

综上所述，TCP 流的传输速率主要与传输文件的大小、传输协议(d,RTO,RTT)和网络使用状况(p)相关。在传输协议各种参数相对固定的情况下，网络使用状况对传输速率的影响十分显著。

2.2 UDP 流和 ICMP 传输速率模型

UDP 流和 ICMP 流从其协议的角度分析，并不保证可靠的传输，其传输速率主要是与发送端系统的发送能力(譬如，局域网中网络适配器的发送能力、ADSL 中上行和下行的可用带宽等等)和上层应用协议的行为(譬如，在一定时间内需要发送的数据量)相关，而与源宿端点之间实际可用带宽以及网络的拥塞状况不存在明显的相关关系。由于系统的发送能力与硬件性能相关，而且除非存在大数据量的传输，否则在一般情况下端系统的发送能力要高于网络中可用带宽，因此分析这两种协议类型的 IP 流需要从高层协议和网络使用者的行为进行分析。

2.2.1 *UDP 流的传输速率模型*

由于 UDP 协议不保证可靠传输，对串路使用状况也不敏感，因此 UDP 流在传输过程中主要是依靠上层协议来控制报文的发送速率。也就是说，UDP 流的传输速率主要受上层协议和用户行为的影响，而且其主要功能是用于实际数据交互而非控制。前期文[6,15]研究认为，在网络中基于 UDP 协议的最主要两种应用类型是实时媒体和 DNS 交互。然而今年来，由于即时通讯工具在国内网络中被普遍使用，而这些即时通讯软件一般都会同时开启 TCP 和 UDP 端口进行通讯(如 MSN 和 QQ 等)，因此在目前 Cernet 系列 TRACE 中相当部分的 UDP 流来自于即时通讯应用。

本文分别对来自不同 TRACE 的 UDP 流传输方式进行抽样统计，抽样流长的阈值为 10 且不为无应答流，抽样数量为 200 个。发现，UDP 流主要的交互模式可以大致分为两种：如图 2 所示。属于前一种交互模式(如上图所示)的 UDP 流中，存在较多源宿双方交互所使用的端口其中之一或者全部为 53，由于 53 为提供 DNS 服务的周知端口，因此可以判断该 UDP 流是由于 DNS 的交互所产生的。后一种交互模式(如图 2 所示)，从微观上看存在突发现象，而从宏观的角度分析，该类型 UDP 流是呈现平稳增长的趋势。由于应用的不同，属于该类型 UDP 流每次突发所发送报文数量也是各不相同的(从抽样所获得的数据来看，每次突发所发送的报文数从 2～16 不等)。而且，即使在同一个 UDP 流中，每次突发产生的报文数量也不一定相同，因此即时通讯和实时流媒体的流

量特征均符合这种交互模式。

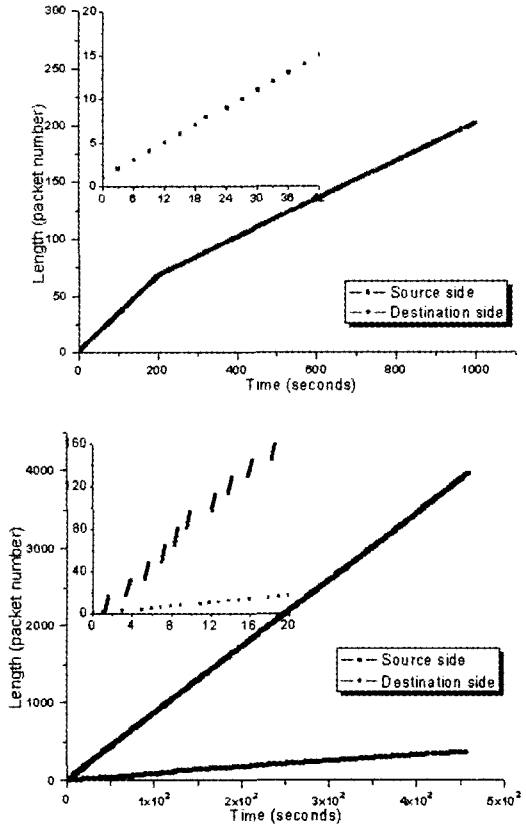

图 2　两种典型 UDP 流的报文到达序列

从源宿双方产生的交互报文比例上看，前者双方产生的报文数量相等，而后者两者数量产生明显的差异，但是每次接收端在接收到报文之后一般会产生一个或若干报文响应发送端，因此在一般情况下，UDP 流的双向报文数为偶数的几率远大于奇数。

从以上分析可以看出，不管是何种交互模式的 UDP 流，源宿端点之间的交互都可以看作由发送阶段和等待阶段这两部分组成：在发送阶段，UDP 报文在短时间内被发送到网络中，这些报文到达后接收方会发送相应应答报文；进入等待阶段，源宿端点均进入静默阶段，没有任何报文的交互，这两个阶段交替进行。

假设 k 为每次交互源端点发送报文数量，$m \leqslant 1$ 为交互时源宿双方双向报文数量比率，N 为 UDP 流源宿端点交互次数，则 UDP 流的流长为

$$N_{pack} = E[k] \cdot (1 + E[m]) * N \tag{11}$$

UDP 流的持续时间可以表示为

$$T = E[T_B] \cdot N + E[T_T] \cdot (N-1) \tag{12}$$

其中 $E[T_B]$ 为每个发送阶段平均需要时间，$E[T_T]$ 为等待阶段平均所需时间。

因此 UDP 流的平均流速率可以表示为突发发送报文数量、双向报文数量比、发送时间和等待阶段持续时间等的函数，由于发送时间（毫秒级）相对于等待时间（秒级）十分短，因此平均流速主要与突发时发送的报文和等待时间相关。

2.2.2　ICMP 流传输速率模型

ICMP 流主要承担网络探测的作用。从流长分析来看，ICMP 流平均流长远小于 10 个报文，因此 ICMP 流的传输速率受网络上层应用（如 ping，Traceroute 等）所驱动，在较短时间发送一系列报文，用于探测对方主机的活动情况。这种探测行为可能是一次或者重复多次，因此从观测的角度分析，绝大部分 ICMP 流的传输报文数为

$$N_{pack} = (E[s] + E[r]) \times n \tag{13}$$

其中 $E[s]$ 为源端主机每次探测发送 ICMP 报文的数量，$E[r]$ 为宿端主机发送 ICMP 报文响应的数量，n 为探测行为重复的次数。

ICMP 流的持续时间为

$$T = E[T_S] \cdot n + E[T_T] \cdot (n-1) \tag{14}$$

其中 $E[T_S]$ 为每次探测平均所需时间，$E[T_T]$ 为两次探测行为之间平均等待时间。

由此 ICMP 流的平均流速主要是由每次探测报文平均发送数量、探测重复次数、平均等待时间以及发送数量等相关因素所决定的。

3　IP 流平均流速分析

在本文中，IP 流的平均流速定义为单位时间内属于特定流的到达的报文数量。对于一个特定的 IP 流而言，平均流速为该 IP 流的流长除以流持续时间。平均流速用于表达 IP 流源宿端点之间报文交互的平均速率，从而反映该 IP 流对网络带宽的占用状况、源宿端点间网络可用性等相关信息。在固定观测点获取的平均流速分布还可以反映网络流量的分布状况、各种网络具体应用协议和用户行为对网络性能的影响。

本节主要从 IP 流平均流速总体分布和不同协议类型 IP 流平均流速等若干不同角度对平均流速进行分析，并寻找影响 IP 流平均流速的若干主要因素。

3.1　数据来源

为反映各种不同时间、不同应用背景和不同负载的大规模高速网络 IP 流分布状况，本文所选取的 TRACE 主要来自于：(1) CERNET 华东北地区网络中心在不同时段获得的 TRACE；(2) 美国互联网研究国家实验室（NLANR）[16] 公开提供的 TRACE。具体的 TRACE 分布概况如表 1 所示。

表 1　不同来源 TRACE 分布概况

TRACE	Duration	Avail_BW	bps(M)	pps(K)
Cernet_a	Nov. 10, 2005 00：40 (20 minutes)	1G * 2 * 3 (6Gb)	1505	271
Cernet_b	Nov. 10, 2005 21：00 (1 hour)	1G * 2 * 3 (6Gb)	3583	701
Abilene I	Aug 14, 2002 09：00 (2 hours)	OC48 (2.5G)	846	138
AbileneIII	Jun 01,2004 19：31 (30 minutes)	OC192 (10Gb)	1822	226
Leipzig-I	Nov 22,2002 20：00 (10 hour)	OC3 (155M)	43	10.41
AucklandII_a	Dec 01,1999 19：25 (24 hours)	100M	0.53	0.22
AucklandII_b	Jan 28, 2000 16：00 (24 hours)	100M	0.76	0.25
AucklandIV_a	Mar 11,2001 12：58 (24 hours)	100M	1.56	0.49
AucklandIV_b	Apr 05, 2001 01：00 (24 hours)	100M	3.20	0.97
CESCA-I	Feb 19,2004 10：40 (1 hour)	1G	560	115

Duration：为选用 TRACE 的起始时间和持续时间，均以当地时间为准；Avail_BW：表示网络设计可用带宽；bps：表明每秒平均传输比特数，也就是 TRACE 采集时实际平均占用带宽；pps：表明每秒平均传输报文数。

CERNET 系列 TRACE 的采集点位于江苏省教育网边界路由到国家主干路由之间；其他 TRACE 均来自美国互联

网研究国家实验室在不同时间对不同网络采集的结果，有关TRACE的详细资料在其网站[6]上均有详细介绍。本文选取了从2001年到2004年不同采集点、不同带宽的8个TRACE作为分析对象，以尽量减小IP流宏观特性的研究由于网络采集点用户偏好所导致的偏差。其中Abilene系列的TRACE来自于从Indianapolis(IPLS)到Cleveland(CLEV)主干网络，带宽较大，应用也相对复杂；Auckland系列的TRACE由于来自于Auckland大学的边界，而且采集的时间也比较早，因此可用带宽和实际平均使用带宽比较小，应用类型也比较单一；Leipzig-I不管是带宽还是数据的采集时间均位于前两者之间，应用也相对于Abilene系列较简单而相对Auckland系列较复杂；而CESCA-I只提供了一个方向的流量，由于其位于Catalan R&D网络的边界，用户特征相对其他网络具有一定的可比性，本文也采用其作为研究的对象。

3.2 IP流平均流速总体分布

本节依然从分析不同TRACE的角度，讨论IP流平均流速总体分布情况，以及IP流平均流速主要受哪些因素的影响。图3分别从流数量和报文数量的角度使用互补累计分布曲线(CCDF)分析了IP流和非TCP流的平均流速总体分布情况。

图3(a)描述了从流数量的角度考察所有IP流平均流速分布状况。从分布分析可以看出，所有TRACE中，IP流在平均流速分布上依然服从重尾分布，但是CERNET系列和Abilene系列等来自于大规模网络的TRACE与Auck系列的TRACE在分布上存在明显的差别，这些差别主要在于前者在平均流速分布中存在明显的突变状况，这些突变主要集中在平均流速为0.33报文/s和0.66报文/s附近。图3(c)描述了报文数量角度考察的IP流平均流速分布状况，通过对比不同TRACE的分布曲线可以看出，大规模网络中主要负载是由流速较快的IP流承担的。譬如：Auck系列TRACE中只有不到40%的报文是由流速在10报文/s的IP流传输的，而在CERNET系列和Abilene系列TRACE中对应比例为60%以上。其中在AbileneI中，这个比例高达84.8%。另外，通过对比图3(a)和(c)可以看出：流数量的突变对报文数量的分布并没有很大的影响，因此可以初步推测流数量突变主要是由短IP流引起的。

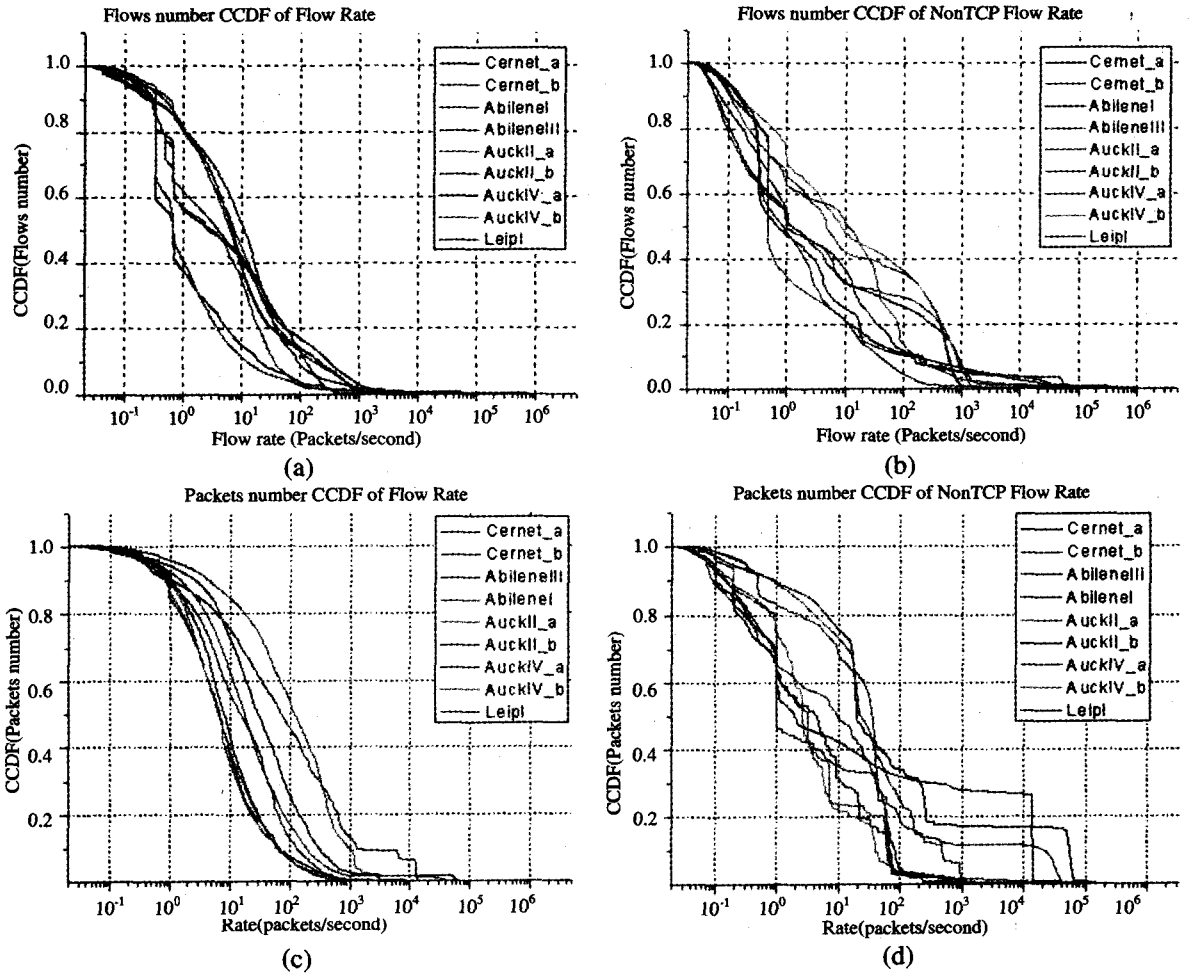

图3 IP流流速的互补累积分布曲线

由于TCP流在IP流中占主要地位，因此有必要考察非TCP流的平均流速分布情况，以及它们对总体分布的影响。图3(b)和(d)分别描述了从流数量和报文数量的角度考察非TCP流平均流速分布状况。在图3(b)中，非TCP流的CCDF在下降速度上明显大于相同类型的IP流，这说明平均流速较慢的流在非TCP流中占较多的份额，而且在平均流速较慢处存在一些流数量突变的状况。譬如，CERNET系列约为0.33报文/s处，LeipI在约为0.50报文/s处，Abilene系列和Auckland系列TRACE在1.00报文/s处等等。在图3(d)Cernet系列和LeipI等TRACE中报文数量分布在平均流速为0.33报文/s和0.50报文/s处虽然也存在一定的突变，但其幅度相对于流数量变化要小得多，因此可以推测在此处发

生突变的IP流平均流长较小;而与在平均流速为1.00报文/s处,Auckland系列TRACE中报文数量的突变幅度明显大于流数量的突变(如AuckII_b的报文数量变化值为19.5%,而对应流数量的变化值只有4.1%),因此推测在此处发生突变的非TCP流平均流长要高于总体平均流长。

通过对比分析,可以看出IP流的平均流速分布状况类似于持续时间分布,从流数量的角度分析存在明显的重尾现象,而从报文数角度分析重尾现象并不显著。因此可以初步推断,从宏观角度分析,流长较大的IP流的平均流速高于流长较小的IP流。在IP流平均流速分布上也存在若干明显的突变显现,而且这些突变存在一定的规律性,需要作进一步分析。

3.3 TCP流平均流速分布

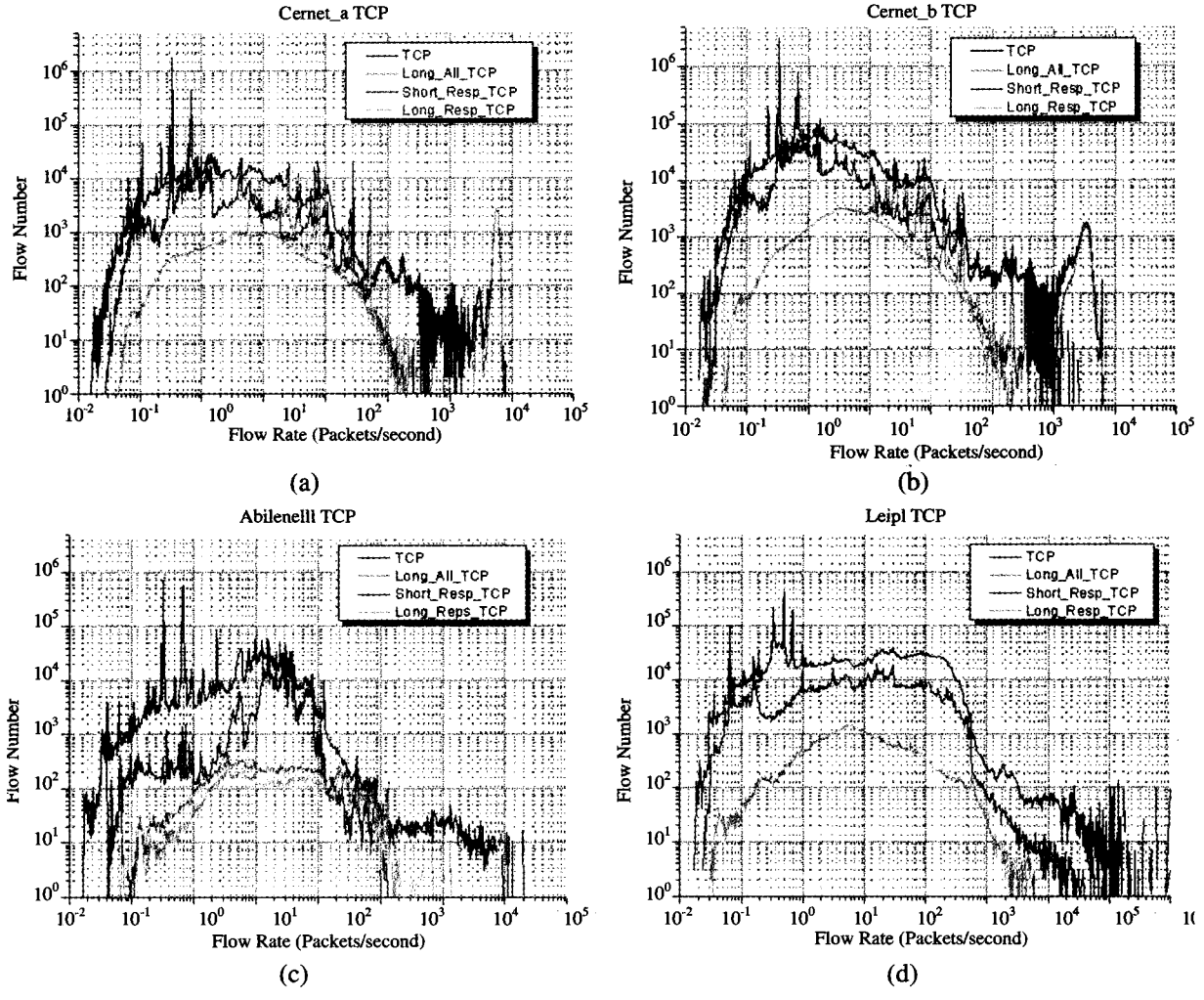

图4 不同类型的TCP流的流速分布

本节将对相应的四组TRACE进行分析,比较它们在平均流速上的异同和分布特点,如图4所示(其中纵轴为实际观测所得的流数量,横轴为流的平均流速)。

本节除了考察这些TRACE中TCP流总体平均流速分布(图中以"TCP"标记)之外,还分别对长TCP流(图中以"Long_All_TCP"标记)、有应答的短TCP流(图中以"Short_Resp_TCP"标记)和有应答的长TCP流(图中以"Long_Reps_TCP"标记)进行了考察。

从总体分布来看,平均流速主要存在以下特点:(1)短TCP流的平均流速分布范围较广,存在较多的流数量突变状况;(2)长TCP流的平均流速分布表现出比较明显的对称分布状况,大量TCP流平均流速集中在某个值附近;(3)除Leipl系列外,"Long_All_TCP"和"Long_Reps_TCP"这两条曲线存在明显的差异,但是导致差异的原因根据TRACE的不同也各不相同:CERNET系列TRACE中,该差异主要是由于路由循环所引起的,因此差异主要集中在流速较快处(如流速大约为80.0报文/s,300.0报文/s等处);而Abilene系列TRACE中差异分布范围较广。由于这些TRACE采集自主干网络,路由不对称可能是导致差异的最主要原因。

在所有TRACE中,平均流速为0.33报文/s处均存在不同程度的TCP流数量突变,其中CERNET系列中该现象比较显著。通过进一步分析发现,在该处流长为3的无应答TCP流占绝大多数。由公式(7)可知,当$RTO=3, k=2$时,平均流速为0.33报文/s。这就说明了这些TCP流主要是由连续三次尝试建立串接但建立串接失败的TCP流组成。

由2.1节中对TCP流不同阶段的流长和流持续时间公式可知,对于TCP流而言,影响其平均流速的主要因素为丢包率p、RTT和RTO。对由于不存在应答而不能建立正常串接的TCP流,其平均流速主要受RTO影响;若能TCP串接能正常建立,则对应的TCP流的平均流速随着丢包率的增加而减小,其原因主要是:由公式(10)可知,对于大多数TCP流而言,RTO对平均流速的影响随着丢包率的增加而增加;而对短TCP流而言,初始RTO值一般远大于RTT值。由于短TCP流中存在大量未能正常建立串接的未应答TCP流,因

此短 TCP 流受 *RTO* 的影响要明显大于长 TCP 流。这一点从平均流速的对比分析也可以看出:从总体上看,短 TCP 流的平均流速要小于长 TCP 流。

在短 TCP 流平均流速分布上,也存在明显的 TRACE 采集地点相关性。譬如,图 4(a)和(b)中来自 CERNET 系列的 TRACE 中"TCP"和"Short_Resp_TCP"在分布上均存在明显的相似性。这些相似性不仅包括总体趋势,而且包括各种突变现象。因此可以初步推测,网络总体 IP 平均流速分布特征是与具体网络相关的。但是长 TCP 流特别是应答长 TCP 流从平均流速分布上分析,并不存在明显测量网络相关性。譬如,图 4 (a)、(b)和(d)中,来自 CERNET 系列的 TRACE 与来自 Leipl 的 TRACE 在"Long_Reps_TCP"这个分布曲线上并不存在明显的差别。

3.4 非 TCP 流平均流速分布

对四组 TRACE 中非 TCP 流的平均流速分布分析如图 5 所示,也分别从四个角度:所有非 TCP 流(图中以"NonTCP"标记)、长非 TCP 流(图中以"Long_All_NonTCP"标记)、有应答的短非 TCP 流(图中以"Short_Resp_NonTCP"标记)和有应答的长 TCP 流(图中以"Long_Reps_NonTCP"标记)来考察。

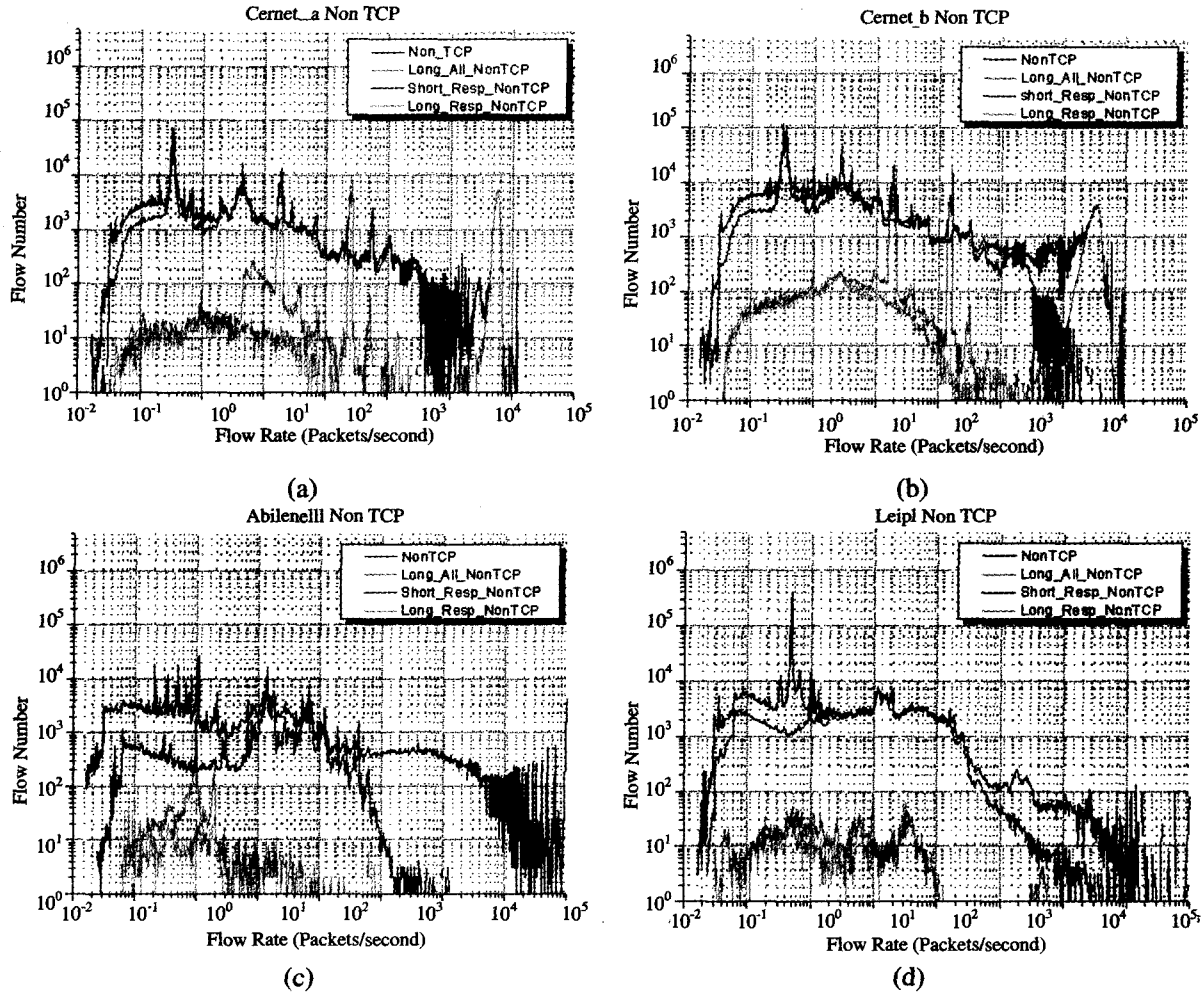

图 5 不同类型 NonTCP 流的流速分析

从平均流速的总体分布来看,非 TCP 流中短流所占的比例明显高于对应 TCP 流中短流,这一点从关于流长实际分布也得到论证。由于短流占绝对多数,因此非 TCP 流的总体分布主要反映对应短流的分布状况。非 TCP 流也存在流数量突变情况,主要集中在 0.33 报文/s、1.00 报文/s、4～4.5 报文/s。通过进一步测量发现,平均速率 0.33 报文/s 的 IP 流中一个端口号为 53 的 UDP 流占绝对多数,这说明这些 IP 流主要是由 DNS 交互产生的;在平均速率为 1.00 报文/s 的 IP 流中,ICMP 流占绝对多数。在通用操作系统(如 Windows 和 Linux)中提供的最常用网络可用性探测工具 Ping 是基于 ICMP 协议,其每隔 1s 发送一个报文,因此如果在接收不到应答的情况下,该工具产生的 ICMP 流从平均流速的角度来看接近于 1.00 报文/s。

从总体来看,本文所分析的所有 TRACE 中,长非 TCP 流的平均流速明显小于对应的长 TCP 流,这种状况并不随着网络状况和用户偏好的变化而变化。由分析可知,只有极小部分 ICMP 流的流长大于 100,因此本文中所描述的长非 TCP 流绝大部分为 UDP 流。在 2.2 节中有关 UDP 流持续时间和平均速率分析可知,UDP 流持续时间主要是由报文突发时间和等待时间这两个部分构成。如果不考虑等待时间,那么 UDP 传输的平均速率只与所需传输报文的大小和发送报文的时间相关(如公式(15)所示),而与串路状况无关。

$$V=V_{pack}/T=(E[k]\cdot(1+E[m])\cdot N)/(E[T_B]\cdot N)$$
$$=E[k]\cdot(1+E[m])/E[T_B] \qquad (15)$$

如果 TCP 流传递相同的数据量,还必须考虑串路的状况,而且 TCP 传输从建立串接到稳定传输还存在慢启动的过程。因此在这种情况下,UDP 流平均速率应高于 TCP 流平均速率,这一点在文[Yilmaz01]中介绍短 TCP 流、长 TCP 流

和 UDP 流三者之间的关系时也有所论述。但是实际观测结果是，UDP 流的平均流速明显低于 TCP 流的平均速率，因此可以推断在绝大多数长 UDP 流中等待时间是不可忽略的。

由于非 TCP 流不存在建立可靠串接和拥塞控制等问题，因此这部分 IP 流的平均速率主要是由高层应用控制，与其所需传输数据量也有密切关系，但受网络层协议和网络状况的影响较小，所以这方面的因素基本可以忽略。

对于短非 TCP 流而言，由于传输报文数量比较少，因此其平均速率主要是由其所包含的等待时间的长短和频率所决定，因此高层应用的特征十分明显，如前文介绍的由应用程序 Ping 产生的无应答 ICMP 流的平均流速均为 1.00 报文/s。长非 TCP 流主要由 UDP 流构成，其传输速率模型在 2.2 节中已经作过详细的介绍，实际测量的结果也基本符合该模型，主要是受高层应用的影响。

4 IP 流平均流速特征

从平均流速分布特征来看，在一定较长的时间范围内，相同采集节点不同时段获得的 TRACE 具有类似的分布，因此推测流速分布的总体趋势与采集节点有较强的相关关系，由此推测网络的基础设施和使用者的行为决定了 IP 流的平均流速。归纳各种 IP 流的平均流速特征如下：

(1) 大部分短 IP 流的平均流速较小，这主要是由于短 IP 流在不同程度上受到网络状况和使用者行为的影响的结果。短 TCP 流平均流速受网络协议、网络运行状况和网络使用者的行为影响，其他非 TCP 协议类型的短流平均流速主要受网络使用者行为的影响；

(2) 受传输协议和高层应用协议的影响，IP 流数量在平均流速分布上存在突变现象。这些数量突变主要集中在短 IP 流处，而长 IP 流的分布几乎不受突发的影响；

(3) 流速较快的 IP 流承担了网络中的绝大部分流量；

(4) 随着网络规模的扩大，平均流速较快的 IP 流所承担网络负载比率也随之扩大。

(5) 在网络中，由于受高层应用的影响，UDP 流的平均流速要小于同样流长的 TCP 流。

总结 本文首先分别通过对 TCP 流、UDP 流和 ICMP 流使用基于协议分析的方法，研究它们在一般情况下的平均流速模型。还针对具有拥塞控制机制的 TCP 流位于不同工作状态的特点，分别建立了不同状态下 TCP 流的平均流速模型，得出了不同类型 IP 流中决定平均流速的主要影响因子。

在讨论 IP 流平均流速分布的影响因子时，需要分别从 TCP 流和非 TCP 流这两个角度分析，其最主要的区别在于前者使用的 TCP 协议对网络环境敏感，能够根据网络所处环境的情况调整发送报文的速率，即所观测到 IP 流的速率；后者因为不能从网络中获取反馈信息，故对网络环境和带宽使用状况不敏感。

影响 TCP 流平均速率的主要因素可以分为使用者行为因素和网络因素这两个方面：(1)使用者行为一般直接反映为网络中各种不同应用层协议所占的比例。从前文文献分析可知，随着网络规模的扩大和网络中应用协议类型的改变，长 IP 流所占比例有增长的趋势，流长的改变导致了网络中 IP 流平均流速和持续时间的改变；(2)网络因素在这里主要指传输层协议对 IP 流平均流速的影响，由于 TCP 协议需要建立可靠的串接以及根据丢包率调整报文的传送速率。

UDP 流和 ICMP 流不受网络环境的影响，所以在传输层层次上对 IP 流平均流速是不存在影响的。从 3.4 节分析可以看出，它们主要受使用者行为，也就是不同应用层协议所占比例的控制。

另外从分析中可以看出，在平均流速的分布上存在一定量的 IP 流流数量突变情况，这些突变的原因主要由两部分构成：(1)传输层协议，大量未能正常建立串接的 TCP 流反映了相同的特征；(2)应用层协议，如由应用程序 Ping 形成的 ICMP 流和由于 DNS 交互形成的 UDP 流等。

综上所述，IP 流平均流速的影响因子主要来自于两个方面：使用者行为和网络行为。其中使用者行为对非 TCP 流(主要指 UDP 流和 ICMP 流)平均流速的影响较大，而 TCP 流的平均流速一般是由这两方面的因素共同决定的。由于本文主要集中于定性和部分定量讨论 IP 流平均流速的影响因子，因此下阶段工作主要围绕本文提出的这些影响因子在实际网络中的定量关系展开。

参 考 文 献

1 Sarvotham S, Riedi R, Baraniuk R. Connection-level Analysis and Modeling of Network Traffic. In: Proc. of ACM 1st Internet Measurement Workshop. San Francisco, 2001. 99～103

2 Sarvotham S, Riedi R, Baraniuk R. Network and user driven alpha-beta on-off source model for network traffic [J]. Computer Networks, 2005, 48: 335～350

3 Brownlee N, Claffy K. Understanding Internet Traffic Streams: Dragonies and Tortoises [J]. IEEE Communications Magazine, 2002, 40(10): 110～117

4 Zhang Y, Breslau L, Paxson V, et al. On the Characteristics and Origins of Internet Flow Rates [A]. In: Proc. of Sigcomm 2002 [C]. USA, 32(4): 309～322

5 Lan K C, Heidemann J. On the correlation of Internet flow characteristics: [Technical Report]. ISI-TR-574 [R]. USC/Information Sciences Institute, July 2003

6 Lan K C, Heidemann J. A measurement study of correlations of Internet flow characteristics [J]. Computer Networks, 2006, 50 (1): 46～62

7 Lu D, Qiao Y, Dinda P, et al. Characterizing and Predicting TCP Throughput on the Wide Area Network. ICDCS, 2005. 414～424

8 Ryu B, Cheney D, Braun H W. Internet Flow Characterization: Adaptive Timeout Strategy and Statistical Modeling [J]. In: Workshop on Passive and Active Measurement (PAM), Apr. 2001

9 Postel J. Transmission Control Protocol. IETF, RFC793. Sept. 1981

10 Cardwell N, Savage S, Anderson T. Modeling TCP latency [A]. In: IEEE INFOCOM [C]. March 2000. 1742～1751

11 Padhye J, Firoiu V, Towsley D, et al. Modeling TCP Throughput: a Simple Model and its Empirical Validation [C]. In: Proc. of the ACM SIGCOMM'98 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication. Canada, 1998. 303～314

12 Li Y, Williamson C. A Hysteresis Model for Web/TCP Transfer Latency. In: Proc. of IEEE/ACM MASCOTS. Volendam, Netherlands, 2004. 167～174

13 Aikat J, Kaur J, Smith D, et al. Variability in TCP Round-trip Times. In: Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement. Miami, 2003. 274～284

14 Shakkottai S, Srikant R, Brownlee N, et al. The RTT Distribution of TCP Flows in the Internet and its Impact on TCP based Flow Control [R]. http://www.caida.org/publications/papers/2004 /tr-2004-02/

15 Lan K C, Heidemann J. Multi-scale Validation of Structural Models of Audio Traffic: [Technical Report]. ISI-TR-544 [R]. Nov 2001

16 http://pma.nlanr.net/[EB/OL]