大规模互联网活动 IP 流分布研究[1]

程光　龚俭　丁伟
（东南大学计算机科学与工程系　南京 210096）
(江苏省计算机网络技术重点实验室　南京 210096)

摘要：随着 Internet 主机数量的持续增长，理解活动 IP 流行为对于网络流量计费、网络安全管理、流量测量、仿真等研究具有重要的指导意义。本文根据 2001 年年初进出 CERNET 华东（北）地区网络的一个月流量，从流量数据宏观角度系统分析了接入网络内部 IP 同外界交换信息的活跃程度。研究分析表明入 IP 流、出 IP 流和 IP 流的统计规律具有重尾分布的属性，并以此建立相应的分布模型及提出评价分布模型的方法。

关键词：统计分析、活动 IP 流、重尾分布、Pareto

Research on the distribution of activated IP flow in the large-scale networks

Cheng Guang　Gong Jian　Ding Wei
(Department of Computer Science & Engineering, Southeast University, Nanjing 210096)
(Key Lab of Computer Network Technology Jiangsu Province, Nanjing, 210096)

Abstract: In recent years, because the number of the computers in the Internet has been increasing, it is important for network accounting, network safe management and network flow measurement to see the behavior of the activated IP flow. In the paper, we have researched almost 30TB traffic trace that all traffic entered into and removed from the CERNET Eastern China (north) regional network in Feb. 2001. And we have analyzed the active degree of IP in the regional network. The research result indicate that the statistics of in-IP flow, out-IP flow and IP flow is heavy-tailed and the ratio of in-IP flow and out-IP flow has the character of log-normal distribution. According to these characters, we make the model of pareto distribution and log-normal distribution and propose the way that evaluate these distribution models.

Keywords: Statistical Analysis, Activated IP Flow, Pareto, Log-Normal Distribution

1. 引言

近几年 Internet 主机一直成指数增加，现在联网主机已经超过 1 亿台[1]，网络用户的行为对 Internet 的发展将起重要的影响，因此研究网络用户行为对于网络规划、网络管理、网络 QoS 服务、网络应用设计等研究都有重要的意义。网络用户的活跃程度可以通过对网络 IP 流量统计研究来发现，理解活动 IP 流统计行为对于网络流量计费、网络安全管理、流量测量等研究均有重要的指导意义。目前，研究活动 IP 流行为已成为网络行为研究的一个重要方向。

国外在研究网络流量统计性质上做了很多工作，并发现了一些重要的流量性质。Paxson[2]对 Internet 进行了历史意义的研究，建立了 5 种常用网络应用流量分布模型；Thompson[3]分别在 1997 年 5 月和 8 月分别持续 24 小时和 7 天监测一条 OC-3 主干上的 Internet 流量，分析流量的分组大小、流的持续时间、协议和应用组成等特性；Mena[4]研究多媒体流量的分布特性；Mah[5]检查 Web 流量日志记录建立了 Web 流量分布经验模型，Crovella[6]发现 Web 文件服从重尾分布；Leland[7]等人对以太网流量进行统计分析发现以太网流量具有自相似特性。

[1]基金项目：本文受国家自然科学基金重点课题“90104031”和国家 863 课题“2001AA112060”资助

这些工作主要从流量组成的微观角度刻画流量特性，本文在总结这些工作的前提下从流量宏观统计角度研究一个广域网内部 IP 同外界 IP 交换信息的统计分布模型，研究成果可以用于对用户行为进行宏观流量监控和分析，制定相应的网络管理政策，满足相应的 QoS 要求。

为了研究活动 IP 流的统计模型，我们在 2001 年年初从 CERNET 华东（北）地区网络中心对 CERNET 华东(北)地区网与 CERNET 主干网交换流量持续进行了 28 天的监测。通过对实测流量进行统计研究发现：活动 IP 流分布服从重尾分布，根据这些特性建立相应的 Pareto 模型。

2. 流量测量

2.1 基本概念

在叙述流量测量之前，先对文中使用的几个重要概念进行定义。

定理：设 S 为所有被关注 IP 地址所构成的集合，ξ_i 为集合 S 中的地址 i 在关注时间内流入（出）的信号总量，为一族独立同分布的随机变量。则，ξ_i 服从重尾分布。

为了证明本文的结论，本文用一个月实测的入流量、出流量和总流量分别进行研究。

本文对活动 IP 的讨论以流为对象。根据研究问题的不同有不同的流定义，Claffy[8]提出了一种用参数的方法对各种粒度下的 Internet 流进行摘要。

定义 1：入 IP 流是指在测量时间内，从接入网络外部到接入网络内部一个活动 IP 的所有报文流量的总和。

定义 2：出 IP 流是指在测量时间内，从接入网络内部的一个活动 IP 到接入网络外部所有报文流量的总和。

定义 3：IP 流是指在测量时间内，接入网络内部的一个活动 IP 进出接入网络所有报文流量的总和。

上面 3 个关于 IP 流定义中的间隔时间均是按照测量时间计算，文章取 2001 年某月为流的间隔时间，在这段时间内到达的分组均属于其 IP 流。3 种流均是以接入网络内部 IP 为集合粒度，不考虑接入网络外部的 IP 地址以及应用层使用服务类型的差别。

值得注意的是文中的活动 IP 不等于一台上网主机，因为一台主机有可能有多块网卡，每个网卡对应一个 IP，或者 IP 地址分配是采用 DHCP 动态地址分配，在不同的时间同一 IP 可能对应不同的主机。另外，文中的活动 IP 是流量的发起方或接受方，因此 Internet 网关和路由器的 IP 不属于活动 IP。

2.2 数据测量

本文中使用的网络流量数据来自于 CERNET 华东(北)地区网络中心对 CERNET 华东(北)地区网与 CERNET 主干网交换流量的监测。CERNET 华东(北)地区网络是 CERNET 全国 8 个地区网络之一，连接江苏、安徽、山东的 140 多所高等院校和研究单位。2000 年下半年起该地区网与 CERNET 主干网的互联信道从 OC－3 升级至 OC－48 后，两网之间的流量以及每天高峰流量和低峰流量的比值都发生了很大变化。我们在 CERNET 华东（北）地区网络中心使用自行研制的高速 IP 网络监测系统对 CERNET 华东(北)地区网与 CERNET 主干网交换流量进行了监测，并取了一个月的监测数据作为本文的工作内容。

表 1：监测期活动 IP 流统计摘要

	IP 流	入 IP 流	出 IP 流
活动 IP 数目	24,089	23,922	21,200
最小长度	100k	100k	100k
最大长度	14,019,897,089k	323,184,949k	13,751,340,856k
平均长度	1,203,012.8k	295,268.3k	1,033,765.7k
中值	19,926k	18,671k	2,161k

在所取的监测期内共观测到 27,415 个活动 IP，总 IP 流流量为 28,979,423,145KB，其中

出 IP 流流量为 21,915,972,485KB，占总流量 75.63％；入 IP 流流量为 7,063,450,660KB，占总流量 24.37％，表 1 为活动 IP 流量大于 100k 的统计摘要。

从表 1 中可以知道最大出 IP 流的流量相当大，达到近 14TB，占总流量的 48％，平均流量在 1GB 左右，而中位值仅 2MB，平均值是中位值的 500 倍。入 IP 流流量的平均值近 300M，中位值不到 19M，平均值是中位值的 16 倍。活动 IP 流流量这种平均值远大于中位值的属性说明活动 IP 流具有重尾属性。

由于流量采集位于地区网络边界上，有部分流量是通过代理服务器转发的，这样代理服务器将校园网内部的活动 IP 的流量屏蔽了，我们得到代理服务器 IP 流，而实际流量发起和接收的内部 IP 并不能发现。另外，文中使用采集的数据只是地区网络内部 IP 同地区网络外部 IP 流量交换的流量，对于地区网络内部 IP 之间交换的流量是不加考虑的。这有利于进行主干活动 IP 流分析统计建模，因为我们只考虑 IP 对 Internet 主干产生的影响，仅在局域网或校园网内部的流量不对主干有任何影响，如果考虑这部分流量将会使分析经验模型偏离主干流量实际情况。这种活动 IP 流统计分布模型有利于网络流量采集和流量计费处理。当然如果我们要研究的研究对象是局域网，那么流量测量位于局域网边界上，这样可以采集所有进出局域网的流量。

3. 统计模型

建立统计分布模型有两种方法：一种是经验分布模型，Tcplib[9,10]对这种模型进行过讨论，这是一个广域网流量的经验模型，它根据 Internet 站点上实测的数据建模；另一种是分析模型，这种模型是用简单的数学描述而不是经验分布，因而更容易进行转换和分析，本文采用统计分布模型的就是分析模型。这一节先讨论重尾分布和对数正态分布的统计分布模型，然后讨论分布模型的统计检验方法。

3．1 统计分布模型

3.1.1 随机变量

文中定义随机变量为活动 IP 流的流量，不同流量的活动 IP 流的概率是不同的，每个活动 IP 流的流量为随机变量的一个实例。一个随机变量的实例是在测量期间，入 IP 流流量或出 IP 流的流量。另一个随机变量的例子是这些流量的对数值。传统上，使用 X 代表一个普遍的随机变量而 x_i 是随机变量 X 的第 i 个实例，文中假设总共有 n 个实例。

3.1.2 概率分布模型

定义随机变量的分布函数为：设有一随机变量 X，它可取任何可能的值 x。现欲求小于或等于某 x 值（即 X≤x）时的概率。显然，其概率依赖于 x，也就是说，P 是 x 的一个函数。使 $F(x)=P(X\le x)$，也就是 F(x)是随机变量 X 小于或等于 x 的概率。

（1） 在分析模型中，我们发现活动 IP 的入流量和出流量比率的分布服从对数正态分布。若 x 为对数正态随机变量，则 X=lgx 的的概率密度函数定义为

$$f(x)=\frac{1}{x\sigma_{\lg x}\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\lg x-\mu_{\lg x}}{\sigma_{\lg x}}\right)^2\right] \qquad (1)$$

式中 $\mu_{\lg x}$ 与 $\sigma^2{}_{\lg x}$ 分别为变量 lgx 分布的均值与方差，其估计值分别为

$$\bar{x}_{\lg x}=\frac{1}{n}\sum_{i=1}^{n}\lg x_i\ ,\quad s^2{}_{\lg x}=\frac{1}{n-1}\sum_{i=1}^{n}(\lg x_i-\mu_{\lg x})^2 \qquad (2)$$

在有限次测定中，可分别用样本均值 $\bar{x}_{\lg x}$ 与样本标准差 $s_{\lg x}$ 来估计总体均值 $\mu_{\lg x}$ 与总体标准差 $\sigma_{\lg x}$。将与取反对数之后，分别得到

$$G = \lg^{-1} \bar{x}_{\lg x}, \quad s_g = \lg^{-1} s_{\lg x} \qquad (3)$$

G 和 s_g 分别称为几何平均值与几何标准差。

（2） 在本文中我们发现活动 IP 流流量大小具有具有重尾属性的分布。设随机变量 X 遵循重尾分布，则：$P[X > x] \sim x^{-\alpha}$，当 $x \to \infty$，$0 < \alpha < 2$，最简单的重尾分布是 Pareto 分布，具有概率密度函数

$$p(x) = \alpha k^{\alpha} x^{-\alpha-1}, \quad \alpha, \quad k>0, \quad x \geqslant k \qquad (3)$$

累计分布函数为：$F(x) = P[X \leq x] = 1 - (k/x)^{\alpha}$ （4）

可以使用 Log-log 补分布图(CD)的方法估计经验测量分布中的 α 值和 k 值，CD 图在 log-log 轴上的补累计函数 $\overline{F(x)} = 1 - F(x) = P[X > x] = (k/x)^{\alpha}$，用这种方式画出重尾分布的图具有属性 $\frac{d \log \overline{F}(x)}{d \log x} \sim -\alpha$，根据相关的尺寸大小，以 log-log 为尺度画出它们的余分布函数图。

$$\log_{10}(1 - F(x)) = \alpha \log_{10} k - \alpha \log_{10} x \qquad (5)$$

由（5）可以知道 $\log_{10}x$ 和 $\log_{10}(1-F(x))$成线性关系。在实际中，我们通过绘制数据集的 CD 图得到 α 的估计值。根据 Pareto 的对数补概率分布函数（5），通过最小二乘法估计出 α 和 $\alpha \log_{10} k$，设 $y_i=\log_{10}x_i$, $z_i=\log_{10}(1-F(x_i))$,可以估计出 α 和 k 的值为:

$$\hat{\alpha} = -\frac{\sum (y_i - \bar{y})(z_i - \bar{z})}{\sum (y_i - \bar{y})^2}, \hat{k} = 10^{-\frac{\bar{z} - \hat{\alpha}\bar{y}}{\hat{\alpha}}} \qquad (6)$$

3.2 数据处理

在我们进行数据分析前需要对数据进行对数转换和剔除一些数据。如果分析的数据在一个方向无限而另一个方向有限，通常使用对数转换来重新表示数据。在本文的数据分析中，为了能说明取值范围很大的数据，本文的许多模型需要对数据进行对数转换。在模型中常采用 $\log_2 x$ 或 lgx 转换。如果随机变量 Y=logX 服从正态分布，那么我们称 X 服从对数正态分布。

使用对数转换处理非负数据时，会遇到数据的值等于 0 的问题，因为对数转换将使数据变为－∞。因此在数据处理之前我们需要将活动 IP 流中流量接近 0 的数据剔除。在本文中我们将流量小于 100K 的活动 IP 流认为是不活跃 IP，不活跃 IP 的流量占总流量的 0.0001％以下，这些 IP 的活跃程度对网络流量没有任何影响，因此，在数据处理中将这些 IP 剔除不会影响模型的结果。

3.3 分布假设检验

我们根据统计资料选择了某一分布函数 F(X)，则统计分布 $F^*(X)$与理论分布 F(X)必有某

些分歧。产生分歧的原因有两种：

1）分布函数选择不适当；

2）分布函数与统计分布模型不矛盾，而分歧是由于观测次数

所谓分布的假设检验就是要制定一个标准来判断分歧的原因是属于前者或是后者。本文中测量分析模型偏差的技术是采用 λ^2 检验法。讨论 λ^2 检验法之前先研究 x^2 检验法。

假设一个有 n 个测量实例的随机变量 Y，使用分布函数 Z 建模，我们将分布函数 Z 成为 N 个箱，每个箱子具有概率 p_i，p_i 为分布函数 Z 符合第 i 个箱子的概率，假设 Y_i 是实际观测 Y 落在第 i 个箱子的个数，则可得 x^2 为 $\chi^2 = \sum_{i=1}^{N} \frac{(Y_i - np_i)^2}{np_i}$，简单地使用 x^2/n 作为 x^2 差异测量。然而使用 x^2/n 作为 x^2 差异测量具有一些问题，一般说来，当计算 x^2 随 Y 的大小和它的标准方差变化时使用最优的 N 值（箱子的数目），但 x^2/n 偏差测量不能用来比较不同 N 值之间的差异。

考虑到 x^2/n 偏差测量不能比较不同 N 值之间的差异，可以使用 λ^2 差异测量代替 x^2/n 差异测量来解决这一问题，因为 λ^2 能用来比较不同 N 值的差异。下面讨论 λ^2 评估随机变量 Y 和模型分布函数 Z 之间的差异。首先假设 $E_i=np_i$ 是第 i 个箱子的期望数，$D_i=Y_i-E_i$ 为第 i 个箱子的偏差，那么定义 $K = \sum_{i=1}^{N} \frac{D_i}{E_i}$，然后定义

$$\hat{\lambda^2} = \frac{X^2 - K - df}{n-1} \qquad (7)$$

其中 df 是计算 X^2 和 K 自由度的数目，在本文中 df=N-1-est，est 是 Y 需要进行估计的参数数目。估计 $\hat{\lambda^2}$ 相应的方差为：

$$\sigma(\hat{\lambda^2}) = [2df + 4n\hat{\lambda^2} + 4n\hat{\lambda^4} + 4T]/n^2 \qquad (8)$$

其中 $T = \sum_{i=1}^{N} [D_i^3 - 2D_iE_i + \frac{5}{2}D_i^2 + \frac{3}{2}(D_i + E_i)]/E_i^2$，$\hat{\lambda^2}$ 的标准差为：$\sigma_\lambda = \sqrt{\sigma(\hat{\lambda^2})}$。

4. 统计分析

使用上文讨论的方法，我们建立活动 IP 流、入 IP 流、出 IP 流以及入 IP 流/出 IP 流比率的分布模型。表 2 摘要了随机变量的统计分析模型的相关特性，“变量”栏中列出建模中使用的随机变量。“模型”栏中列出用来描述随机变量分布的分析模型，其中 4 个是 Pareto 模型，1 个是 Log_{10}-normal 对数正态模型，这些模型是有限制的，第一个仅描述流量大小在 40%~99%的 IP 流，第二个模型描述 IP 流流量 99％以上的 IP 流，第三个模型描述入 IP 流流在 80％以上的入 IP 流，第四个模型描述出 IP 流流量在 10％以上的入 IP 流，最后一个模型是描述入/出 IP 流比率大于 1 的区间。“参数”栏中给出模型中使用的参数，参数表中使用“≈”代替“＝”。“λ^2 范围”栏中给出使用第二中偏差估计方法的标识。“α 90％置信区间”栏给出了前 4 种模型 α 的 90％置信水平，α 置信区间的差别是由于三种流不同的特性引起的，α 的大小反映了随机变量的尾部特性。“R^2”栏是前进行 4 种模型的 CD 曲线进行线性拟合程度的标志，越接近 1 表明随机变量具有良好的线性关系。下面我们将详细讨论表 2

中的模型。

表 2： 活动 IP 经验统计分析模型摘要

原型	变量	模型	参数	λ^2	α 90%置信区间	R^2
IP 流	流量	Pareto,40%~99%	α ≈0.476 k≈12.08	0.05	(0.43,0.51)	0.973
IP 流	流量	Pareto,99%~100%	α ≈0.787 k≈15464.2	0.04	(0.757,0.817)	0.975
入 IP 流	流量	Pareto,80%~100%	α ≈0.913 k≈22,582	0.02	(0.883,0.943)	0.986
出 IP 流	流量	Pareto, 10%~100%	α ≈0.635 k≈445.47	0.03	(0.605,0.665)	0.984

4.1 活动 IP 流分布模型

活动 IP 流流量前 500 个 IP 流占总流量的 90%。当然目前重点是研究前 20 位的活动 IP 的情况，这 20 个 IP 流量占总流量的 80%，即不到千分之一的活动 IP 的流量占总流量的 80%，因此，这些 IP 的活动行为对总的网络流量行为有重要影响。而出 IP 流流量分布曲线尾部较重，前 8 个出 IP 流量占总输出流量的 95%，前 500 个 IP 输出流量占总流量的 99%。相对来说，入 IP 流分布较为平均，排名前 15 的 IP 仅占总输入总流量的 20%。

通过对总流量、输出流量和输入流量的分析可以看出：输出流量占总流量的 75.6%，输入流量占 24.4%，因此，地区网络是输出的大于输入流量。由于输出流量集中在几个服务器，因而流量分布极度不均，前 8 个（即万分之三）IP 流量占总输出流量的 95%，前 500 个（2%）IP 输出流量占总流量的 99%。而入 IP 流流量分布较均，输入流量占前 15 的 IP 仅占总流量的 20%，前 14000 个（50%）IP 占总输入流量的 99%，这是因为入 IP 流流量主要是由客户机完成，流量较大的输入流量大多是代理服务器。

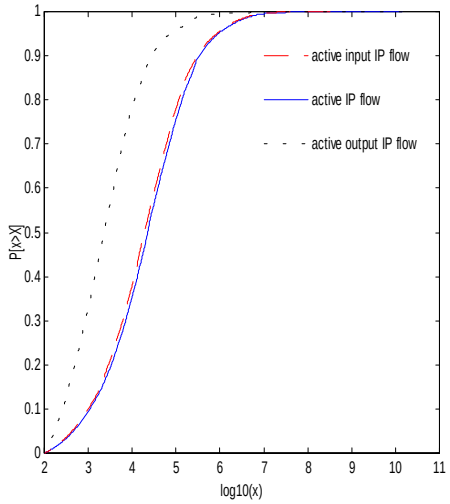

图 1 活动 IP 流分布曲线图

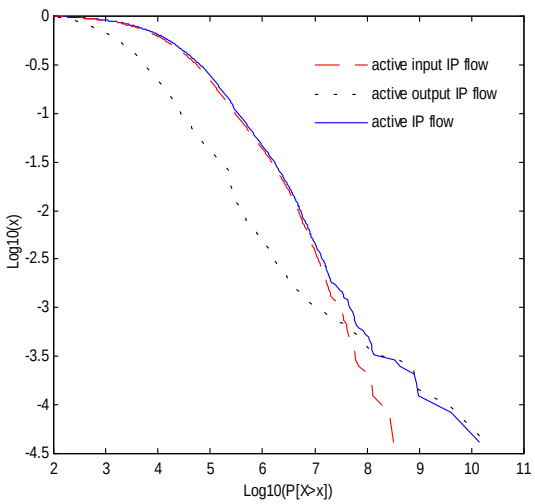

图 2 活动 IP 流 CD 图

图 1 为活动 IP 流对数分布曲线图，从图中可以看出入 IP 流和 IP 流的分布曲线较为接近，而出 IP 流分布曲线则分离较远。图 2 为活动 IP 流的 CD 曲线图，图中可以看出 log(X)<7 左右 IP 流的 CD 曲线同入 IP 流非常接近，而在 log(X)>8 的 IP 流的 CD 曲线同出 IP 流相当接近，从 CD 中可知 log(X)>8 的 IP 仅占总数的约 0.03%，根据表 1 可以知道，出 IP 流的流量集中在少数的几个大出 IP 流，重尾曲线较大。而入 IP 流分布较为均匀，重尾曲线较小。在图 2 中，出 IP 流和入 IP 流的曲线交点约在 log(x)＝7.6、log(1-P[X<=x])=-3.2，即 x=40,000,000，P[X<=x]=99.93%。

图 2 中，当入 IP 流大于 10,000KB 时，入 IP 流 CD 曲线基本呈直线，这说明入 IP 流大

于 10,000KB 时呈重尾分布，将 log10(x)>4 的出 IP 流的 CD 曲线进行最小方差线性拟合，可以得到得到（R^2=0.986）斜率为-0.913，常数 3.975，使用式（6）可以计算出相应的$\hat{\alpha} = 0.913$，$\hat{k}$ =22,582.85。R^2=0.986 说明直线拟合效果明显。

在图 2 中，当出 IP 流大于 1,000KB 时，出 IP 流 CD 曲线基本呈直线，这说明出 IP 流大于 1,000KB 时呈重尾分布，将 log10(x)>3 的入 IP 流的 CD 曲线进行最小方差线性拟合，可以得到得到（R^2=0.984）斜率为-0.635，常数 1.682，使用式（6）可以计算出相应的$\hat{\alpha} = 0.635$，$\hat{k}$ =445.47。R^2=0.984 说明直线拟合效果明显。

下面我们考察 IP 流的分布模型，从图 2 中可以知道，IP 流 CD 是由两部分组成，在 log10(x)=8 是转折点，log10(x)大于 4 小于 8 的流大小由于受入 IP 流的大小影响较大，斜率接近入 IP 流曲线，而 log10(x)大于 8 的 IP 流由于受出 IP 流影响加大，因而斜率接近出 IP 流曲线。当 log10(x)>8 时，通过拟合 IP 流 CD 曲线，可得拟合直线 R^2=0.973，斜率为-0.476，常数为 0.515，使用式（6）式可以计算出$\hat{\alpha} = 0.476$，$\hat{k}$ =12.08。当 4<log10(x)<8 时，拟合直线为 R^2=0.975，斜率为 0.787，常数为 3.297，对应的$\hat{\alpha} = 0.787$，$\hat{k}$ =15464.17。

5．结论

本文采用统计分布模型对活动 IP 流进行统计分析，建立相应的 IP 流、入 IP 流、出 IP 流以及入 IP 流和出 IP 流的统计分析分布模型，并且分析了评价分析模型同实测记录偏差比较的统计方法，模型的建立使我们能更深入地理解活动 IP 的特性，进而能从宏观上理解用户的行为特性，研究成果对于网络行为学的研究与实践及网络容量规划和网络管理具有重要意义。

从我们的研究可以发现活动 IP 具有以下规律：1）活动 IP 流呈现 Pareto 重尾分布，10％的活动 IP 流的流量占总流量的 90％以上；2）出 IP 流同入 IP 流相比极度不均，出 IP 流分布函数重尾于入 IP 流，0.03％的出 IP 流占总出 IP 流流量的 95％以上，而 0.03％的入 IP 流仅占总入 IP 流流量的 15％；3）入 IP 流同出 IP 流的比值总体上呈对数正态分布，95％的 IP 是入流量大于出流量，75％入 IP 流流量和出 IP 流流量比值在 10～100 之间，分布中心在 25 左右，也就是说在被观测的互联网中 95％以上的活动 IP 应该是客户机，小于 5％的活动 IP 是信息资源服务器。文章分析的模型和方法能准确地反映接入网络内部 IP 同外界交换信息的活跃程度。虽然我们分析的 CERNET 活动 IP 流的统计分布情况，应该可以推测，其它如中国电信等其它大规模都具有类似分布特性，整个 Internet 流量行为也应该具有以下重要属性，即：输出流量集中在极少数活动 IP 中，而输入流量分布却较为分散，大规模网络只有极少数 IP 是属于文件服务器。

本文的研究成果对于网络安全监控有重要意义。目前国际网络主干速度由 OC48 增加到 OC192，由于硬件和软件的限制，为了网络管理、计费、问题诊断或其它目的而监控网络越来越困难，如：目前的网卡没有捕获网络流量 OC48 速度的能力、主机的 I/O 限制、硬盘和 RAM 的限制等。根据本文的发现的活动 IP 流重尾分布的特点，设置相应的过滤机制，将属于重尾 IP 的流量过滤，将大大减轻测量负担。如：本文的研究 CERNET 地区网络进入 CERNET 主干的流量中前 8 个 IP 流流量总和占总出 IP 流流量的 75％，如果将这 8 个 IP 地址的流量过滤，那么需要测量的日最高流量降为原来的 1/4，同时也大大减轻需要进行安全分析的负担。通过采用这种过滤方案，目前能进行 OC12 速度测量的网卡不用进行任何改进

就可以直接用于 OC48 速度测量，具有 OC48 速度的网卡可以直接用于 OC192 速度测量。在主干速度升级期间，不致于由于网络速度过快而无法进行测量。

参 考 文 献:

[1] Number of Hosts advertised in the DNS, Internet Domain Survey, July 2000 http://www.isc.org/ds/WWW-200007/index.html.

[2] Vern Paxson, Empirically-Derived Analytic Models of Wide-Area TCP Connections, IEEE/ACM Transactions on Networking, August 1994.

[3] Kevin Thompson, Gregory J. Miller, and Rick Wilder, Wide-Area Internet Traffic Patterns and Characteristics, IEEE Network, November/December 1997.

[4] Art Mena, John Heidemann, An Empirical Study of Real Audio Traffic, In Proceedings of the IEEE Infocom, p. 101-110. Tel-Aviv, Israel, March 2000

[5] B. Mah, An Empirical Model of HTTP Network Traffic, In Proceedings of the IEEE Infocom, April 1997.

[6] Mark E. Crovella, Murad S. Taqqu, Azer Bestavros, Heavy-Tailed Probability Distributions in the World Wide Web, In A Practical Guide To Heavy Tails, pages 3--26. Chapman & Hall, New York, 1998.

[7] W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson, On the Self-Similar Nature of Ethernet Traffic, IEEE/ACM Transaction on Networking, vol.2, Feb. 1994.

[8] K.C. Claffy, H.-W. Braun, and G.C. Polyzos. Parameterizable methodology for internet traffic flow profiling. IEEE Journal on Selected Areas in Communications, 13(8):1481--1494, October 1995.

[9] Peter B. Danzig, Sugih Jamin, tcplib: A Library TCP Internetwork Traffic Characteristics,

[10] Ramon Caceres, Peter B. Danzig, Sugih Jamin, Danny J. Mitzel, characteristcs of wide-area TCP/IP Conversations, In ACM SIGCOMM, 1991.

[11] George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel 著, 顾岚 译, Times Series Analysis Forecasting and Control, 中国统计出版社, 1997.

[12] 杨位钦, 顾岚 著, 时间序列分析与动态数据建模, 北京理工大学出版社, 1988

作者简介：程光，博士研究生，主要研究方向为网络管理、网络行为学。龚俭，教授、博士生导师，主要研究方向为网络安全、网络管理、网络体系结构