第 34 卷第 1 期 通 信 学 报 Vol.34 No. 1
2013 年 1 月 Journal on Communications January 2013
doi:10.3969/j.issn.1000-436x.2013.01.003

UDP 流量对 TCP 往返延迟的影响

朱海婷，丁伟，缪丽华，龚俭
(东南大学 计算机科学与工程学院 江苏省计算机网络技术重点实验室，江苏 南京 211189)

摘 要：提出一个面向单条链路的 TCP 综合传输性能测度 *R*，分析其与同一时间粒度内链路的占用带宽和 UDP 流量比例间的关系，使用中国和美国 2 条主干信道的实测数据进行了检验。结果表明 *R* 测度可表示为以占用带宽和 UDP 流量比例为参数的正态分布的随机过程。随后进行的 CERNET 链路 *R* 测度正态分布均值参数的拟合分析具有很高的可接受水平，拟合的结果给出的量化关系可用于流量的公平性评价。

关键词：TCP 性能；UDP 流量模型；流量公平性；往返时延
中图分类号：TP393 **文献标识码：**A **文章编号：**1000-436X(2013)01-0019-11

Effect of UDP traffic on TCP's round-trip delay

ZHU Hai-ting, DING Wei, MIAO Li-hua, GONG Jian
(School of Computer Science & Engineering, Southeast University, Key Laboratory of Computer Network Technology in Jiangsu, Nanjing 211189, China)

Abstract: A single link-oriented TCP transmission performance metric called "*R*" was proposed. The relationship between metric *R* and two indicators (the utilized link bandwidth and the UDP traffic proportion) on the same link was analyzed, and the experiment results on passive measurement datasets from two bidirectional network links separately located in USA and China were laid out. The results show that the metric *R* can be expressed as a normally distributed stochastic process with distinct parameters decided by the utilized bandwidth and the UDP traffic proportion. Further, by using fitting methods on local datasets, a concrete function, which was acceptable at a high confidence level, was given to describe the relationship between the mean value of *R*'s normal distribution and the two indicators mentioned before. This work offers a reference to the study of network fairness.

Key words: TCP performance; UDP traffic model; network fairness; round-trip time

1 引言

TCP 和 UDP 是自互联网诞生之日起就开始使用的传输层协议。在过去的 20 多年里对新的传输层协议的研究和讨论曾经成为相关领域的研究热点，并有以 IETF 标准形式提出的 TCP 友好速率控制协议 TFRCP（TCP-friendly rate control protocol）[1] 和用户数据报拥塞控制协议 DCCP（datagram congestion control protocol，IANA 协议号 33）[2]为代表的成果得到公认。然而，根据自 2005 年以来对 CERNET（China education & research network）一条 OC192 省网边界链路的持续观测[3]，除了 2%左右的流量是和路由有关（ICMP 等）或封装报文外，其余 98%全部由使用 TCP 和 UDP 协议（IANA 协议号分别为 6 和 17）的报文构成。协议号为 33 的报文数量为零。美国的互联网数据分析机构 CAIDA（cooperative association for Internet data analysis）发布的数据同样证明了这一点[4,5]。这说明在过去的近 20 年里，虽然使用不同应用协议的应用软件已发展到数以千计，但它们几乎全部使用 TCP 和 UDP 这 2 个传统的传输层协议。AKHSHABIS 在文献[6]中认为造成这个现象的原因是 TCP 和

收稿日期：2012-05-18；修回日期：2012-11-26
基金项目：国家重点基础研究发展计划（“973”计划）基金资助项目（2009CB320505）；国家科技攻关计划基金资助项目（2008BAH37B04）
Foundation Items: The National Basic Research Program of China (973 Program) (2009CB320505); The National Key Technology Research and Develop Program of China (2008BAH37B04)

UDP 在功能上互补，它们利用先入优势，抑制了所有功能上与其有重叠的其他协议。据此，互联网中传输层协议在未来相当长的时间内仍将由 TCP 和 UDP 主导，而且是二者共存的局面。

在所观测的链路上，UDP 流量比例从 2005 年起逐年上升，至 2009 年后逐步稳定在 50%左右，表 1 给出了关于 UDP 流量比例以及其他几个主要流量测度的变化数据。文献[7]根据中国国内近期某运营商主干网 12h 的观察数据认为，从所有的角度看 UDP 流量已占用了一半左右的网络资源，这与我们的观测结果基本吻合。在国际上，CAIDA 在文献[4]中发布的数据表明 UDP 流仅在随机高端口传输小数据量报文，在整体字节数和报文数上 TCP 仍旧占绝对优势。文献[5]提到 1998~2008 年 10 年中 UDP 与 TCP 报文比例在 5%~20%之间变化，但没有呈现出稳步增长的态势，且观察到的挪威、瑞典、日本和德国以及英国等部分地区的数据也表现出类似特性。表 2 是 CAIDA 网站[4]公布的数据与时间相近的上述 CERNET 链路数据关于 UDP 和 TCP 比例测度的对比，从中可以看出彼此间存在较大的差别。文献[8]通过一个高精度的 P2P 流媒体识别方法对上述 CERNET 链路 2010 年的 2 条长度分别为 1h 的数据进行的流量识别结果表明 PPStream、PPLive、QQLive 和 UUsee 4 款国产 P2P 点播软件所产生的 UDP 流量按报文数计算占到全部 UDP 流量的 20%或总流量的 10%左右。另外还有迅雷、腾讯 QQ 等知名国产软件也部分使用 UDP 作为传输层协议，这些软件在中国的广泛使用，应该是造成中国互联网链路上 UDP 流量比例偏高的原因。

从一般的角度来看，当带宽资源紧张导致拥塞发生时，使用 TCP 的应用可以很快地感知到拥塞的存在并自动放缓发送节奏，但 UDP 不具这样的能力，这样 TCP 让出的带宽会被 UDP 占用。UDP 比例越高对 TCP 流量产生抑制就越大，越不利于拥塞

表 1 TCP 和 UDP 流量比例的变化

日期	TCP 比例/%		UDP 比例/%		UDP/TCP 比例/%		TCP + UDP 比例/%	
	in	out	in	out	in	out	in	out
Nov.10/2005	94.76	94.70	4.91	5.04	0.05	0.05	99.67	99.74
Dec.31/2006	78.73	84.94	21.16	15.03	0.27	0.18	99.89	99.97
Oct.13/2007	77.23	76.01	22.72	23.91	0.29	0.31	99.95	99.92
Mar.16/2009	50.80	58.89	47.86	39.79	0.94	0.68	98.66	98.68
Jun.15/2010	50.80	58.04	47.24	41.10	0.93	0.71	98.04	99.14

表 2 CAIDA 数据和 CERNET 数据 UDP 流量比例对比

数据	日期	时间	方向	总流量		UDP/TCP	
				分组数×10^6	字节数/GB	分组数	字节数
CAIDA equinix-chicago OC192	Mar.31/2009	13:00~14:00	dirA	623	550	0.13	0.03
			dirB	2 858	1 709	0.24	0.09
	Mar.25/2010	13:00~14:00	dirA	208	185	0.16	0.08
			dirB	3 012	2 763	0.20	0.07
	Mar.24/2011	13:00~14:00	dirA	512	232	0.17	0.17
			dirB	1 726	1 804	0.14	0.06
CERNET* jiangsu-backbone OC192	Mar.16/2009	14:00~15:00	in	411	303	0.68	0.57
			out	390	197	0.94	1.15
	Mar.02/2010	14:00~15:00	in	931	643	0.57	0.34
			out	844	443	0.75	0.73
	Mar.08/2011	14:00~15:00	in	680	454	0.92	0.48
			out	549	226	1.02	1.24

*CERNET 的数据进行了四分之一流抽样。

的缓解。这样，在高 UDP 流量的链路上，对 TCP 和 UDP 采用同样的拥塞控制机制对 TCP 流量来说是不公平的。目前，面向异质流网络拥塞控制的公平性研究主要围绕保护正当行为流的问题展开，研究的热点是对用户公平的主动队列管理（AQM, active queue management）算法[9~11]。而面向协议间拥塞控制公平性的策略的研究和相应的 AQM 算法目前相应并不多见。

网络流量模型的研究是拥塞控制的基础工作。虽然网络上的流量可以简单而清晰地划分为 TCP 流量、UDP 流量和其他流量，但在机理模型方面，TCP 协议流的模型已经有比较成熟且获得公认的结果[12]，而 UDP 协议流由于其简单和开放的开环结构，使得上层应用可以采用个性化的控制模式，因此很难给出统一的模型，这应该是构造面向传输层协议的 AQM 算法的难点之一。

实际上，拥塞发生信道上的每个 TCP 连接都能感受到 UDP 的流量比例对拥塞缓解过程的影响，TCP 的应答机制使得这个影响会体现在所有涉及到的 TCP 连接所经过的所有链路上，同时直接或间接相邻链路间的 UDP 流量比例和占用带宽情况也会彼此关联。这说明即使在没有发生拥塞的链路上进行观测，也能在一定程度感受到 TCP 传输性能的变化。本文的研究工作将基于这个基本思路展开，研究单一链路的使用带宽和 UDP 流量比例与同一链路上的 TCP 传输的综合性能间的关系。

本文研究工作通过提出一个可在不同时间粒度条件下均能获得的可以综合反映链路上所有 TCP 连接传输性能的测度——*R* 测度展开。利用 *R* 测度的构成特点获得它的正态分布随机过程模型，然后用 CERNET 和 CAIDA 的实测数据对这个模型进行了检验。随后的方差分析表明 UDP 流量比例和占用带宽是影响 *R* 测度的主要因素。在此基础上，通过对正态分布随机过程均值的拟合，获得 CERNET 链路的 *R* 测度与 UDP 流量比例和占用带宽间的解析关系，卡方检验表明这个关系具有很高的可接受水平。

2　相关工作

对 TCP 与 UDP 二者间存在不公平带宽竞争这个问题的研究，最早出现在文献[13]中。FLOYD S 等通过 NS2 仿真说明了对固定数目的 TCP 流和 UDP 流，在单点瓶颈情况下 UDP 流会侵占 TCP 流的带宽。文献[14]的研究从拥塞控制的角度展开，显示 UDP 流量的存在不仅仅是简单地占用了 TCP 的可用带宽，而是从本质上改变了系统的动态特性。文献通过使用分形图表明了非线性现象的存在，并随着 RED 的不同取值而改变；用 NS2 仿真数据表明，UDP 流量（constant bit rate 模型下 5Mbit/s 和 10Mbit/s 带宽情况）的增加会使得 RED 中可能产生的分形和混沌临界点不易到达。文献[15]对网络中异构应用的公平性进行了分析，其工作思路源于参考文献[16]中的效用函数方法，后者的贡献在于将运筹学中规划问题的一些数学概念和方法引入到带宽分配中来。文献[15]将文献[16]中统一描述的流量分为了弹性流和非弹性流并分别使用不同的效用函数，通过分布式接入控制的方法来达到公平性的目的，最后采用 NS2 仿真方法验证其方法的有效性。文献[17]中的基本模型完全建立在 TCP 连接的基础上，利用这个模型细致分析了主动队列管理算法的有效性，在理论上证明了此类算法存在从畅通态到拥塞态到瘫痪态的相变过程，并给出了相变临界点与系统参数的显式关系。在后续的研究中使用仿真表明 UDP 流量会明显影响到畅通链路可以接纳的 TCP 流数量。其他相关工作均围绕提出新的传输层协议这个目标持续展开[1, 2, 18]。

上述研究结果可以归纳为：UDP 流量会对 TCP 流量产生负面影响；基于 RED 的拥塞控制机制在高 UDP 流量比例的环境下存在风险；所有的研究结果均采用 NS2 仿真作为主要手段。

现有的研究工作存在以下 2 个方面不足，其一是目前包括 NS2 在内的所有的仿真平台均采用 CBR (constant bit rate)或者 VBR (variable bit rate)模型来刻画 UDP 流量，即将所有 UDP 流量视为单向的匀速或变速流量，这与目前网络中的实际情况有着很大的差异。根据笔者的实际观察，除多播应用外，几乎其他所有使用 UDP 的应用在实际运行时均产生双向的 UDP 流，而且 2 个方向的流速没有规律。另外，文献[19]的研究表明，目前使用中的 TCP 协议会出现不符合协议规定的现象，主要体现在初始拥塞窗口（ICW）、不规则的重传和流时钟等方面。因此用仿真进行的与流量（特别是 UDP 流量）有关的研究，结果的准确性是有待考证的。其二，没有能被普遍接受的 UDP 流量比例与 TCP 传输性能间的关系模型。

AQM 算法方面，经典的 AQM 算法是 RED,

但它在公平性方面的处理不理想。首先在公平性上对 RED 做出改进的是 FRED[9](flow-RED)，但因其要求有单流信息，可扩展性差。此后的研究热点转移到不需单流信息的公平 AQM 算法，先后提出的一些有影响力的算法包括：RED-PD[20]、SFB[10]、CHOKe[11] 、 SAC[21](self adjustable CHOKe) 、BLACK[22]、ESD[23]、WARD[24]和 DCN[25]等 。但这些算法均没有涉及如何处理 TCP 和 UDP 间的公平性。

3 综合 TCP 传输性能测度

根据 RFC2330[26]中的定义，网络性能相关测度的 2 个要素是明确的含义和具体的计算方法。本节从实际层面来分析可用的网络性能参数来代表 TCP 的传输性能，提出建立在往返迟延这一基本测度上的面向链路的新测度。

3.1 测度的定义

重新选择的测度以往返时延（RTT, round-trip time）为基础。理由是：其一根据 IPPM 框架 RFC 2330[26]的观点，RTT 是衡量网络状态的一个重要指标，对于单个用户来说，端到端的 RTT 值可以表征网络的通畅程度，它可以作为网络性能分析的依据：其次，RTT 与 TCP 关系密切，TCP Reno 的流量模型[20]为

$$T(pr, rtt) = \frac{MK}{rtt\sqrt{pr}} \tag{1}$$

其中，T 是单条 TCP Reno 流的吞吐量，K 是常数，M 是最大报文长度，pr 是分组丢失率，rtt 是往返时延，这个公式表明往返时延与 TCP 的吞吐量成反比。此外，基于被动测量数据计算 TCP 的 RTT 有许多成熟的算法[27~29]。这些算法的基本原理都是基于 TCP 的应答机制获得包括端系统处理时间在内的每个 TCP 连接的 RTT 值。

影响一条 TCP 连接的 RTT 值因素可能有多种，任何一个 TCP 连接的 RTT 不能代表链路上其他 TCP 连接的性能，本文用链路在一定时间粒度内 TCP 连接的平均 RTT 作为该链路上 TCP 传输性能的测度。具体定义如下。

定义 1 设有链路 L 和长度为 Δt 时间片 ts，在 ts 内经过 L 的所有 TCP 流集合为 $TF = \{tf_1, tf_2, \text{L}\}$，对 $\forall i$，设 tf_i 在时间片 ts 内的交互中的平均往返延迟为 r_i，则称

$$R(ts) = \sum_{tf_i \in TF} r_i / |TF| \tag{2}$$

为时间片 ts 上该链路 L 的综合 TCP 传输性能测度，简称 R 测度。在不同的时间片上的 R 测度值用 $R(ts_1)$、$R(ts_2)$ 表示，简写为 R_1、R_2 的形式。$R(ts)$ 可以看作以时间片 ts 为参数的随机过程。

3.2 测度的分析

性质 1 $R(ts)$ 是面向网络状态的随机过程。

虽然从定义表面的形式上看，$R(ts)$ 是以时间片 ts 为参数的随机过程，但决定 $R(ts)$ 特征的并不是其所在的时间粒度的位置，而是由在该时间粒度内可能影响 r_i 的网络测度或状态参数。这意味着在网络状态相似条件下的 R_1、R_2 具有相同分布。

性质 2 如 ts 内 L 上所有 TCP 流的平均往返延迟 r_i 是相互独立随机变量，则 $R(ts)$ 是正态分布的随机过程将在 ts 内 L 上所有的 TCP 流 $TF = \{tf_1, tf_2, \text{L}\}$ 根据源地址和宿地址空间划分成 $M \times N$ 个不相交的子集 $TF_1 \sim TF_{M \times N}$，其中 M 为源地址空间的大小，N 为宿地址空间的大小。划分的原则是每个子集中的 tf_i 具有相似的网络距离，这样可以认为同一子集 tf_i 中的 r_i 服从相同的分布 $f_j(t)$。

设 $\forall tf \in TF_j$ 的概率为 p_j，则 $\forall tf \in TF$ 的平均往返延迟的概率密度函数为 $f(t) = \sum_{j=1}^{nm} p_j f_j(t)$，根据性质的假设条件和中心极限定理，$R(ts)$ 服从正态分布。据此，式(2)可以表示为

$$R(ts) \text{ : } N(a(s), s(s)) = \frac{1}{\sqrt{2\pi s(s)}} \text{e}^{-\frac{(t-a(s))^2}{2s(s)}} \tag{3}$$

其中，s 为 ts 在时间粒度的网络状态。

对于性质 2 需要说明的一点是：文献[30]通过在爱立信公司网络中获得的往返时延的变化序列进行分析，认为在其 RTT 值的分布可以很好地近似为一个截断的正态分布，并给出了实验验证。虽然都涉及到正态分布，但区别在于性质 2 中的 R 测度是在单一测量点对多个 TCP 连接测量所获得的 RTT 值计算出的平均值作为随机变量，而文献[30]的结论是认为在特定条件（相邻 5~7 路由跳数的局域网间）下，以对单个 TCP 连接测量获得的 RTT 值或用 ping 获得的单个往返时延作为取值的随机变量服从正态分布。

4 基于实测数据的检验

本节用来源于 CAIDA 和 CERNET 的 2 条链路

的实测 Trace 数据，对上述 *R* 测度的正态分布特性进行检验。本节首先给出实验数据的具体信息，接着解决了验证工作需要解决 2 个基本问题，即如何定义和获得网络状态及如何确定并获得测度计算所需要 TCP 连接上的往返时延、时间片大小这 2 个参数，最后给出了新测度性质的验证结果和分析。

4.1　分析数据

用于分析的 IP Trace 分别来自于 CERNET 江苏省网边界 jiangsu-backbone 链路[3]（以下简称本地数据）和从 CAIDA[31]下载的 equinix-chicago 链路（以下简称 CAIDA 数据），两者均是 OC-192 主干链路。本地数据的覆盖范围超过 150 所高校，分配的 IP 地址数量近 5 000 个 C 类地址。

选定使用的分析数据集的具体情况如表 3 和表 4 所示。本地数据的 Trace（数据集）总时长为 32h，分为 4 组，每组 4 条，共 16 条。CAIDA 数据的 Trace 总时长为 9h，分为 9 条，每条 1h。每条数据均由 2 个不同方向链路的数据组成，本地数据分为 in 和 out 2 个方向，CAIDA 数据分为 dirA 和 dirB 2 个方向。

表 3　本地数据分析情况

编号	日期	时间	总流量/GB		UDP/TCP 字节数		UDP/TCP 分组数	
			in	out	in	out	in	out
1~1	Nov.17/2009	14:00~16:00	687	453	0.66	1.35	1.26	0.90
1~2	Dec.22/2009	14:00~16:00	690	464	0.61	1.52	0.87	0.69
1~3	Mar.9/2010	14:00~16:00	665	428	0.37	0.86	0.84	0.64
1~4	Mar.23/2010	14:00~16:00	639	459	0.35	0.99	0.86	0.65
2~1	Dec.25/2009	14:00~16:00	685	413	0.65	1.07	1.09	0.77
2~2	Dec.29/2009	14:00~16:00	737	421	0.54	0.87	0.87	0.69
2~3	Jan.5/2010	14:00~16:00	718	444	0.62	1.20	1.14	0.83
2~4	Jun.11/2010	14:00~16:00	695	389	0.39	1.11	0.91	0.70
3~1	Nov.14/2009	20:00~22:00	610	411	1.05	1.81	1.64	1.33
3~2	Jan.12/2010	20:00~22:00	541	436	0.78	1.51	1.47	1.11
3~3	Apr.13/2010	20:00~22:00	585	486	0.50	0.90	1.01	0.78
3~4	May18/2010	20:00~22:00	640	423	0.48	1.07	1.04	0.76
4~1	Jun.15/2010	10:00~12:00	456	245	0.36	0.95	0.76	0.59
4~2	Sep.11/2010	10:00~12:00	477	308	0.30	0.93	0.80	0.58
4~3	Dec.17/2009	10:00~12:00	523	342	0.56	1.10	0.98	0.78
4~4	Mar.18/2010	10:00~12:00	534	364	0.37	0.95	0.81	0.60

表 4　CAIDA 数据分析情况

编号	日期	日间	总流量/GB		UDP/TCP 字节数		UDP/TCP 分组数	
			dirA	dirB	dirA	dirB	dirA	dirB
1	Jan.21/2010	13:00~14:00	260	2 079	0.16	0.04	0.28	0.16
2	Feb.25/2010	13:00~14:00	146	1 329	0.15	0.06	0.32	0.19
3	Mar.25/2010	13:00~14:00	172	2 574	0.08	0.07	0.16	0.20
4	Apr.14/2010	13:00~14:00	203	3 014	0.14	0.08	0.25	0.23
5	Aug.19/2010	13:00~14:00	210	1 800	0.14	0.11	0.18	0.28
6	Sep.16/2010	13:00~14:00	202	2 620	0.10	0.09	0.14	0.23
7	Feb.17/2011	13:00~14:00	139	1 610	0.33	0.18	0.23	0.34
8	Mar.24/2011	13:00~14:00	216	1 680	0.17	0.06	0.17	0.14
9	Apr.13/2011	13:00~14:00	202	1 074	0.24	0.21	0.21	0.36

4.2 网络状态

由于能够使用的信息来源仅限于实测 Trace，因此刻画网络状态的测度必须从 Trace 提供的信息中获得。首先考虑的网络状态测度是使用带宽，因为这是最基本的网络性能测度，其次选择了 UDP 流量比例，这是因为根据引言中的讨论这个测度能对 TCP 的传输性能产生影响，也是本文最为关注的问题。

定义 2 时间片 ts 的网络状态是由带宽属性 p 和 UDP 比例属性 $pu\%$ 构成的二元组。取值分别定义如下。

$p(ts)$：时间片 ts 中所包含的报文总数。

$pu\%(ts)$：时间片 ts 中所包含的使用 UDP 协议的报文总数除以 $p(ts)$ 所得百分比。

选择报文作为处理对象是因为它是路由器进行转发和拥塞控制的基本单位。

选择这 2 个属性的另一个原因是这 2 个属性与 TCP 的往返时延一样，可以通过链路间的相邻关系传递。下面用图 1 中的一个简单结构进行相关说明。假设图 1 中的 L_0 是 Trace 的采集链路，它与 2 条后继链路 L_1 和 L_2 的相关程度为 a_{01} 和 a_{02}，$a_{01}+a_{02}=1$。这样 L_0 上时间片 ts 内的 n 个 TCP 连接中的 $a_{01}n$ 个会将 L_1 作为下一跳链路。当 L_1 发生拥塞时，L_0 上的这部分 TCP 的往返时延同样会受到影响，这个影响会直接反映到基于 L_0 计算的 R 测度中，即使 L_0 的状态完全没有发生拥塞。另一方面，在一般情况下，相邻信道间的 UDP 流量比例和占用带宽这个属性也会彼此关联。以此类推，如果发生拥塞的链路是 L_3，L_0 依然能够通过 R 测度感觉到变化，L_0 与 L_3 上的报文 p 属性和 UDP 报文百分比 $pu\%$ 也会有一定程度的关联。

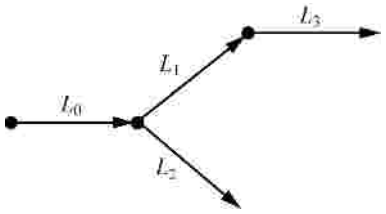


图 1 链路拓扑

这样式(3)可以写为

$$R(ts):N(a(p,pu\%),s(p,pu\%))=\frac{1}{\sqrt{2\pi s(p,pu\%)}}\mathrm{e}^{\frac{(t-a(p,pu\%))^2}{2s(p,pu\%)}} \tag{4}$$

4.3 样本参数

首先是选择可以用于本文工作条件下计算 TCP 连接上往返时延 r_i 的算法，即被动测量数据上的往返时延估计算法。经过比对，最终确定选择使用 SYN-ACK[29]估测和 PRE 算法[27]来作为本文单条 TCP 流 RTT 值的估计算法。具体思想为：先用 SYN-ACK 方法估测出连接初始建立时的 RTT 值作为 PRE 算法的参考值，使用 PRE 算法来获得 RTT 的估测值。

其次是确定时间粒度大小 Δt。选择的时间粒度内需要有足够多的满足上述往返时延 r_i 计算条件的 TCP 连接的数量，这要求 Δt 不能太小。从另一方面，为了使流量的突发性得到保证，时间粒度不宜过长。通过对 4 组数据中数据量最小的 local_out 的流量不同时间粒度内可用于 R 测度计算的 TCP 连接数量和网络流量变异系数（CV）的变化情况的分析和综合考虑，最终取 Δt =5s 作为实验的时间片大小。

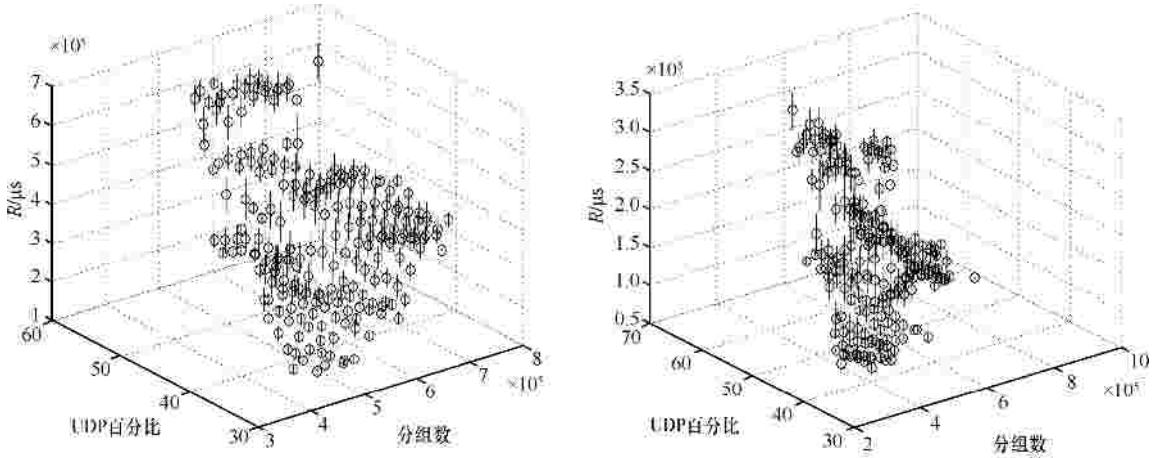
4.4 检验方法和结果

本节工作的目标是基于上述实测 Trace 获得样本，用统计分析的方法验证式(4)的成立情况。式(4)是一个以$(p,\ pu\%)$为参数的正态分布的随机过程，根据 R 测度的宽带属性 p 和 UDP 比例属性 $pu\%$，该二元组取值相近的时间片上的 R 测度应该服从相同的正态分布。进一步，如果将一个时间片上的三元组$(R,\ p,\ pu\%)$作为一个样本观测值，以$(p,\ pu\%)$近似为条件对所有样本进行划分，则划分后同一子集中的 R 测度样本观测值应呈正态分布。本节的验证工作以此为思路进行。

首先是样本观测值的计算。由于时间片大小为 5s，32h 的本地数据可以划分成 $S1$=23 040 个时间片，9h CAIDA 数据的样本数量为 $S2$=6 480。根据定义 2 和 r_i 的计算方法，分别计算每个样本在 2 个不同方向上的$(R,\ p,\ pu\%)$。计算结果如图 2 和图 3 所示。由于本地数据样本数量较大，图 2 中的点是周围一簇点的三元组上的均值，竖线表示这簇值相对于该均值点取值 R 的标准差。从图 2 和图 3 中可以比较直观地看出，无论是本地数据还是 CAIDA 数据，R 都会随 p 和 $pu\%$ 的升高而增大。

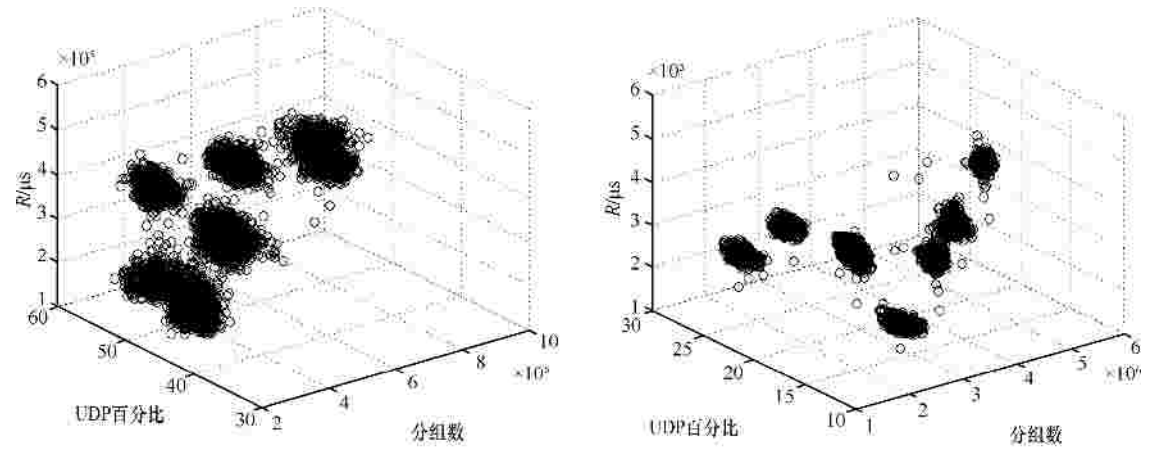
在获得所有样本观测值后，对每条链路按 2 个不同方向，分别用下面的步骤完成正态分布检验。

step1 M 等分该链路上报文属性 p 取值范围，N 等分 UDP 报文百分比属性 $pu\%$ 的取值范围，据此将 R 测度划分成 $M\times N$ 个子集。



(a) 本地数据(方向in)观测结果　　(b) 本地数据(方向out)观测结果

图 2　本地数据的 $(R, p, pu\%)$ 关系



(a) CAIDA(方向dirA)观测结果　　(b) CAIDA(方向dirB)观测结果

图 3　CAIDA 数据的 $(R, p, pu\%)$ 关系

表 5　　正态分布检验结果

$M\times N$	CAIDA dirA			CAIDA dirB			Local in			Local out		
	TN	*AN*	*AN/TN*	*TN*	*AN*	*AN/TN*	*TN*	*AN*	*AN/TN*	*TN*	*AN*	*AN/TN*
5×5	16	4	25.00%	14	5	35.71%	18	6	33.33%	18	4	22.22%
10×10	44	26	59.09%	29	19	65.52%	57	24	42.11%	57	23	40.35%
20×20	107	84	78.50%	56	46	82.14%	164	86	52.44%	163	72	44.17%
30×30	163	143	87.73%	82	65	79.27%	303	203	67.00%	305	163	53.44%
40×40	205	188	91.71%	120	105	87.50%	436	315	72.25%	474	298	62.87%
50×50	224	209	93.30%	145	134	92.41%	532	419	78.76%	606	423	69.80%
60×60	218	202	92.66%	173	159	91.91%	594	489	82.32%	692	532	76.88%
70×70	207	198	95.65%	187	175	93.58%	622	518	83.28%	737	623	84.53%
80×80	151	146	96.69%	209	196	93.78%	641	543	84.71%	749	666	88.92%

注：*TN* 是满足检验条件的子集数量，*AN* 是能够通过正态分布检验的子集的数量。

step2 对满足 MATLAB 中 Lillilifors 检验条件的子集进行正态分布检验，这个条件是样本数量小于 1 000 并大于 10。

4 组数据的检验结果如表 5 所示，其中流量最小的 CAIDA dirA 观测值在 50×50 处获得了最多的检验子集，并在该点上获得 93.3%的接受比例。而另外 3 组数据均在最大的 80×80 处获得最多的检验子集和最高的接受比例。在上述点上，*TN* 中所包含的样本数量占样本总数（*S*1 或 *S*2）的比例如表 6 所示。

表 6 被检测样本比例

CAIDA dirA	CAIDA dirB	Local in	Local out
86.35%	92.53%	85.72%	87.72%

表 5 呈现的检验结果说明式(4)的合理性，其具体的含义是同一时间粒度内的占用带宽属性和 UDP 比例属性可以基本决定 *R* 测度的分布特征，这同时说明即使在没有发生拥塞的链路上，UDP 流量比例也会对同一链路上传输的 TCP 流的传输性能产生影响。另外，表 5 中 CAIDA 数据呈现出更高的正态分布接受水平可能与其样本的时间来源相对单一（均为 13:00~14:00）有关，为了进一步验证这个问题，对本地数据中的第一组单独进行了上述实验，结果如表 7 所示，这个结果也符合上述的设想，这说明式(3)中的网络状态可以进一步将这个因素考虑在内。

表 7 第一组本地数据单独检验结果

M×*N*	*AN*/*TN*/%
5×5	62.50
10×10	77.50
20×20	84.21
30×30	94.08
50×50	97.06

5 分析结果的应用

以上分析表明 *R* 在特定的网络状态下呈正态分布，参数均值 a 和方差 s 与网络状态有关。用同一时间粒度内的 *p* 和 *pu*%作为网络状态测度具有一定的合理性。相对 *R*、*p* 和 *pu*%的获取要简单很多，因此上述研究结果可以用这 2 个测度对 *R* 进行估计，从而获得能够反映同一链路上 TCP 综合传输性能的 *R* 测度。

5.1 网络状态测度对 *R* 的显著性分析

如果将 *R*(*ts*) 看成分析对象，而将 *p* 和 *pu*%看成影响因素，由于分析对象在不同影响因素条件下满足正态分布条件（如表 5 所示），通过显著性分析确定 *p* 和 *pu*%对 *R* 的影响程度。

使用 MATLAB 中多因素方差分析方法 anovan，置信度为默认 95%下，流量大小和 UDP 百分比划分成 80 个等级（其他划分结果均类似）的结果表明 4 组数据均显示 *p*、*pu*%是影响 *R* 的显著性影响因素。

对 *R* 的估计可以直接基于表 5 的结果进行，对一条具体链路 *L* 上的估价方法如下。

step1 选择 *AN* 数量最大的 *M*×*N* （CAIDA dirA 为 50×50，其他 3 条为 80×80），设该划分下所有的 *AN* 构成的集合为 $(an_1, an_2, \text{L}) \ \forall an_i \in AN$，$an_i = (p_{\min}, p_{\max}, pu\%_{\min}, pu\%_{\max}, a, s)$，其中 a、s 为所接受的正态分布检验使用的假设均值和均方差。

step2 对任意时间粒度 *ts*，计算相应时间粒度的 $(p(ts), pu\%(ts))$，若 $(p(ts), pu\%(ts)) \in an_i$，则 *R* 的正态分布均值估计值为 $an_i.a$。

根据正态分布的特性，该方法的精度由 $an_i.s$ 决定，其值落入 $an_i.a \pm an_i.s$ 内为 65.26%，在 $an_i.a \pm 2 \times an_i.s$ 内为 95.44%。

5.2 本地数据均值拟合

用上述方法进行 *R* 值估计所存在的缺陷是明显的。首先，对 *AN* 以外的区间无法给出估计值；其次，所有的估计值为离散的点，很难确定具体的阈值点。为此，本文将用样本数量较大的本地数据，通过拟合，寻找 *R* 与 *p* 和 *pu*%的函数关系，这实际上是对正态分布均值的拟合，即 $a = f(p, pu\%)$。如果这个函数关系存在且合理性得到验证，则可以用其方便地获得 *R* 值的估计，如果该函数是连续性的，则还可以获得性能拐点。

为了提出合适的拟合公式，对图 2 在分组数方向和 UDP 百分比方向分别选取若干切面。图 4 给出了图 2(a)中 UDP 百分比在 36%附近，测度 *R* 与报文数 *p* 的二维关系截面图，图 5 给出了 *p*=5.2×10^5 附近，测度 *R* 与 UDP 报文百分比 *pu*%的二维关系横切面图。

通过观察各个轴上的切面图，*R* 与 *p* 属性呈现接近线性的关系，而与 *pu*%呈现类似指数的关系，据此提出拟合公式

$$a = c_1 p(1 + c_2 \mathrm{e}^{c_3 pu\%}) \tag{5}$$

选取正态符合度较好的表 5 中 80×80 划分方式得到的 801（in）和 762（out）个符合正态分布区域，采用 MATLAB 中的启发式算法 lsqcurvefit 方法，各参数的拟合结果在表 8 中显示。

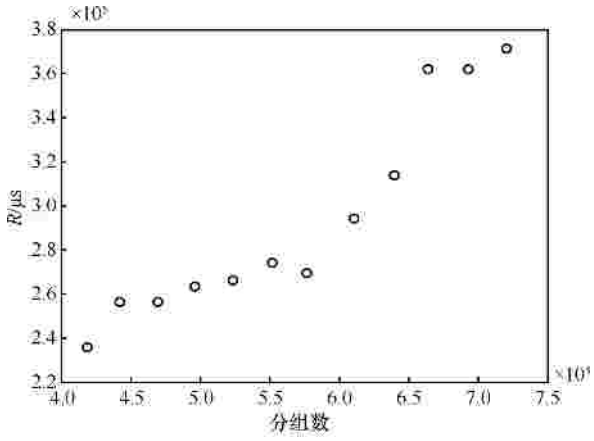


图 4　pu%=36%附近本地数据 in 方向 R 与 p 的关系

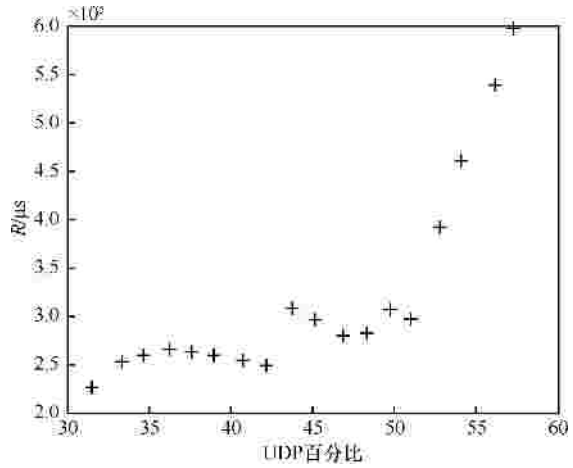


图 5　p=5.2×10⁵ 附近本地数据 in 方向 R 与 pu%的关系

表 8　　拟合参数

数据集	Local (in)	Local (out)
c_1	0.348 1	0.218 4
c_2	0.035 4	3.500 3×10⁻⁵
c_3	6.614 2	16.508 0

5.3　拟合精度分析

选择 2010 年 4 月 13 日 15:00~16:00 的数据（该组数据不是表 3 中的数据），对上述拟合公式从 2 个角度进行拟合效果的分析。一个是拟合误差分析，另一个是正态分布假设检验。

拟合误差是对式(5)进行 R 测度估计准确程度的简单评价。使用的误差公式是 $\frac{|\hat{a}-a|}{\hat{a}}$，$\hat{a}$ 是实测 R 测度值，a 为使用在观察点上网络状态估计其分布的均值 $a=f(p,pu\%)$，因为测度 R 正态分布的特性，其分布中的观察值与估计均值本身就存在一定概率下的误差，根据误差公式的所有观察点上的 R 测度值与其对应分布的估计均值的误差累计分布图 CDF 如图 6 所示。该分析只作为估计精度的一个参考值。

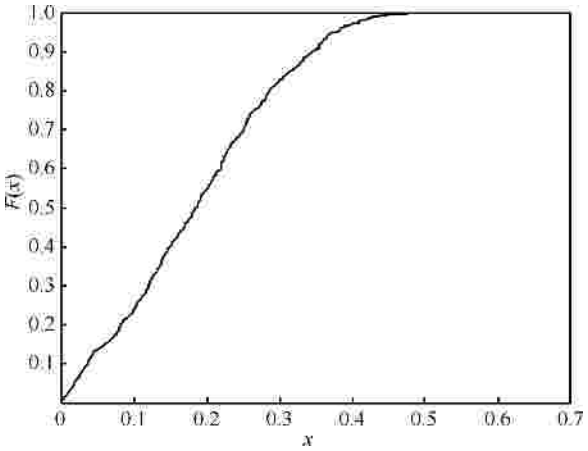


图 6　误差 CDF

正态分布假设检验。该检验将式(5)作为正态分布均值，将方差 s 看成常量，用卡方分析检验相似网络条件下的 R 测度观测值是否符合提出的正态分布假设。具体的操作步骤如下。

step1　以 Δt =5s 划分上述区间获得时间片并计算各时间片被检验数据的网络状态 p、pu%和 R 测度值。

step2　M 等分报文数 p 的取值范围，N 等分 UDP 报文百分比 pu%的取值范围，根据网络状态的取值将 $R(ts)$ 划分成 $M\times N$ 个子集。

step3　对于满足卡方检验条件的任一子集 i（落入样本数量大于 20），用 p 和 pu%的中间值代入式(5)获得其上的正态分布均值估计 a_i。

step4　用卡方方法检验子集 i 中的 $R(ts)$ 样本观测值是否服从正态分布 $N(a_i,s)$，其中，s 取值在 in 方向为 5.832 1×10⁴，out 方向为 2.784 8×10⁴，单位为 MS。

按以上方法完成检验的结果如表 9 所示，表中各符号与表 5 中意义相同，in 和 out 方向分别在 13×13 和 12×12 处具有最大的 TN。这样的检验结果一方面说明正态分布假设的成立，以及用 p 和 pu%来代替网络状态的合理性，另一方面也表明拟合公式(5)的可接受性。

表 9 卡方检验结果

$M\times N$	Local in		Local out	
	TN	*AN*	*TN*	*AN*
5×5	9	4	8	5
10×10	13	10	13	11
12×12	11	10	13	13
13×13	14	14	12	12
15×15	6	6	12	12

5.4 拐点分析

面向性能拐点的分析基于上节拟合结果。需要说明的是这里的性能拐点并不是单纯数学意义上的拐点，而是指曲线的变化程度超过阈值的点。上述拟合函数对 $pu\%$ 的偏导为

$$\partial a / \partial pu\% = c_1 pc_2 c_3 \mathrm{e}^{c_3 pu\%} \tag{6}$$

式(4)是一个一致连续的单调递增函数，这意味着拟合函数切线斜率是连续增加的。取 $pu\%$ 每增加 1%，R 测度变化超过 Δr 为临界点 $pu\%^*$。则对应有

$$pu\%^* = (\ln \Delta r - \ln(c_1 pc_2 c_3)) / (c_3 \Delta pu\%) \tag{7}$$

取 Δr =5ms，选择不同的占用带宽条件，根据式(7)计算出的本地信道 R 测度的性能拐点情况如表 10 所示。拐点的位置会随占用带宽的增大而降低。

表 10 在 Δr =5ms 时的性能拐点 $pu\%^*$

方向	4×10^5	5×10^5	6×10^5	7×10^5
Local in	41.27	37.89	35.14	32.81
Local out	55.74	54.38	53.28	52.35

6 结束语

以 TCP/IP 为核心的互联网体系结构在规模和应用两方面的开放性，使得互联网已经成为一个超大规模的复杂系统，网络使用者行为的随机性还会进一步增加这种复杂性。对这样的系统寻求可以实用的解析模型是非常困难的。统计分析作为一门经典而实用的数学工具，在医学、经济学等同样面向复杂系统的领域里有着非常广泛的应用，基于统计分析所获得结果被广泛接受，说明方法的科学性。

本文的研究工作以统计检验为主要工具，分析的对象为一个面向单链路的 TCP 综合传输性能测度 R，分析数据来源于中国和美国的 2 条链路的实测数据。获得的主要结果包括：R 测度可以表示为以占用带宽和 UDP 流量比例为参数的正态分布的随机过程；这说明即使在没有发生拥塞的链路上，UDP 流量比例也会对综合 TCP 传输性能产生影响；给出了基于占用带宽和 UDP 流量比例对同一时段上的 R 测度的估计方法，其中对大样本条件的本地数据的拟合分析具有很高的准确性和可接受水平。

参考文献：

[1] RFC 5348: TCP Friendly Rate Control (TFRC): Protocol S ecification [S].

[2] RFC 5622: Profile for DCCP Congestion Control ID 4: TCP-Friendly Rate Control for Small Packets (TFRC-SP) [S].

[3] http://iptas.edu.cn, 2011[EB/OL].

[4] http://www.caida.org/research/traffic-analysis/tcpudpratio/,2010[EB/OL].

[5] LEE D, CARPENTER B E, BROWNLEE N. Observations of UDP to TCP ratio and port numbers[A]. Proc Int Conf on Internet Monitoring and Protection (ICIMP) [C]. 2010.

[6] AKHSHABI S, DOVROLIS C. The evolution of layered protocol stacks leads to an hourglass-shaped architecture[A]. SIGCCOM'11[C]. Toronto, Ontario, Canada, 2011.

[7] 张艺濒,张志斌,赵咏等.TCP 与 UDP 网络流量对比分析研究[J].计算机应用研究, 2010, 27(6):2192-2197.
ZHANG Y B, ZHANG Z B, ZHAO Y, *et al*. Comparative analysis on TCP and UDP network traffic[J]. Application Research of Computers, 2010, 27(6): 2192-2197.

[8] WANG G, DING W, DONG S. Classification of UDP traffic from P2P applications[A]. ICON2011, the 17th IEEE International Conference on Network[C]. Singapore, 2011. 252-257.

[9] LIN D, MORRIS R. Dynam ICS of random early detection[A]. Proceedings of ACM SIGCOMM 1997[C]. Cannes, France, 1997. 127-137.

[10] FENG W, KANDLUR D, SAHA D, *et al*. Stochastic fair blue: a queue management algorithm for enforcing fairness[A]. Proc IEEE INFOCOM 2001[C]. Anchorage, Alaska, 2001. 1520-1529.

[11] PAN R, PRABHAKAR B, PSOUNIS K. CHOKe: a stateless active queue management scheme for approximating fair bandwidth allocation[A]. Proc of IEEE INFOCOM 2000[C]. Tel Aviv, Israel, 2000. 942-951.

[12] PADHYE J, FIROIU V, TOWSLEY D, *et al*. Modeling TCP Throughput: a Simple Model and Its Empirical Validation[R]. UMASS-CS-TR-1998-08.

[13] FLOYD S, FALL K. Promoting the use of end-to-end congestion control in the internet[J]. IEEE/ ACM Transactions on ng, 1999, 7 (4):458 - 472.

[14] LA R J, RANJAN P, ABED E H. Nonlinear dynamics of mixed TCP and UDP traffic under RED[A]. Proc of SIGCOMM[C]. Pittsburgh, PA, USA, 2002.

[15] CHIU D M, TAM A S W. Network fairness for heterogeneou lications[A]. Proc of the 1st ACM SIGCOMM Asia Workshop[C]. Beijing, China, 2005.

[16] KELLY P, MAULLOO A K, TAN D K H. Rate control in communication networks: shadow prices, proportional fairness and stability[J]. Journal of the Operational Research Society, 1998, (49):237-252.

[17] 樊华,李理,袁坚等.互联网流量控制的朗之万模型及相变分析 [J]. 物理学报, 2009, 58(11):7507-7513.

FAN H, LI L, YUAN J, *et al*. Langevin model of the flow control in the internet and its phase transition analysis[J]. Acta Phys Sin, 2009, 58(11):7507-7513.

[18] RAI I A, ABDUL S. Towards end-host-based identification of competing protocols against TCP in a bottleneck link[A]. nnals of Telecommunications[C]. 2011. 59-77.

[19] QIAN F, GERBER A, MAO Z M, *et al*. TCP revisited: a fresh look at TCP in the wild[A]. IMC'09[C]. Chicago, Illinois, USA, 2009.

[20] MAHAJAN R, FLOYD S, WETHERALL D. Controlling High Bandwidth Flows at the Congested Router[R]. T R-01-001.AT & T Center for Internet Research at ICSI (ACIRI), 2001.

[21] JIANG Y, HAMDI M, LIU J. Self adjustable CHOKe: an active queue management algorithm for congestion control and fair bandwidth allocation[A]. Proc ISCC 03[C]. 2003. 1018-1025.

[22] CHATRANON G, LABRADOR M A, BANERJEE S. BLACK: detection and preferential dropping of high bandwidth unres onsive flows[A]. Proc of ICC 03[C]. 2003. 664-668.

[23] **汤德佑**, **骆嘉伟**, **张大方等**. **一种提高稳定性和公平性的主动队列管理机制**[J]. **计算机研究与发展**, 2005, 42(7) : 1136-1142.
TANG D Y, LUO J W, ZHANG D F, *et al*. Active queue management improving the stability and fairness[J]. Journal of Computer Research and Development, 2005, 42(7):1136-1142.

[24] HO C Y, CHAN Y C, CHEN Y C. A TCP-friendly stateless AQM scheme for fair bandwidth allocation[A]. Proceedings of the Joint International Conference on Autonomic and Autonomous System and International Conference on Networking and Services (ICAS/ICNS 2005)[C]. 2005. 14-19.

[25] LE L, AIKAT J, JEFFAY K, *et al*. Differential congestion notification: taming the elephants[A]. Proceedings of the 12th IEEE I ternational Conference on Network Protocols, ICNP 2004[C]. Berlin, Germ any, 2004. 118-128.

[26] PAXSON V, ALMES G, MAHDAVI J, *et al*. RFC 2330: Framework for IP Performance Metrics[S]. 1998.

[27] **张轶博**,**雷振明**. **一种被动式** RTT **测量算法**[J]. **北京邮电大学学报**, 2004,27(5):85-89.
ZHANG Y B, LEI Z M. A passive RTT estimate algorithm for TCP[J]. Journal of Beijing University of Posts and Telecommunications, 2004, 27(5): 85-89.

[28] LANCE R, FROMMER I. Round-trip time inference via passive monitoring [A]. SIGMETRICS, 2005[C]. 2005.

[29] JIANG H, DOVROLIS C. passive estimation of TCP round-trip times[A]. SIGCOMM2002[C]. Pittsburgh, PA, USA, 2002.

[30] ELTETO T, MOLNAR S. On the distribution of round-trip delays in TCP/IP networks[A]. Proceedings of the 24th Conference on Local Computer Networks[C]. 1999. 172-181.

[31] http://www.caida.org/data/monitors/passive-equinix-chicago.xml,2012 [EB/OL].

作者简介：

朱海婷（1983-），女，江苏如皋人，东南大学博士生，主要研究方向为网络管理和网络测量。

丁伟（1962-），女，江苏南京人，东南大学教授、博士生导师，主要研究方向为网络测量和网络行为学。

缪丽华（1987-），女，江苏如东人，东南大学博士生，主要研究方向为网络测量。

龚俭（1957-），男，上海人，东南大学教授、博士生导师，主要研究方向为网络管理和网络安全。