# 自相似流量关键参数分析*)

谭献海　黎燕敏　潘启敬　金炜东
(西南交通大学信息科学与技术学院　成都 610031)

**摘　要**　大量的研究结果表明，网络流量过程普遍存在着自相似和长相关特性，自相似和长相关特性对网络性能具有重要的影响。目前绝大部分研究都集中在 Hurst 系数的估计及其性能影响上，这是不全面的。本文深入研究影响网络性能的自相似流量关键参数，通过仿真分析 Hurst 系数和方差系数对网络性能的影响，表明 Hurst 系数和方差系数对网络性能均有重要的影响。分析了方差对网络性能影响的原因，研究了 $C_\gamma$ 与方差之间的关系及其计算方法，给出了基于 IDC 的复合分形更新过程参数的估计算法，分析了分形开始时间对网络性能的影响。

**关键词**　网络流量，自相似，分形布朗运动，复合分形更新过程，计数离差系数

## Analysis of Key Parameters of Self-similar Traffic

TAN Xian-Hai　LI Yan-Min　PAN Qi-Jing　JIN Wei-Dong
(School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031)

**Abstract**　There is mounting experimental evidence that network traffic processes exhibit ubiquitous properties of self-similarity and long-range dependence (LRD), and self-similarity and long-range dependence have great impact on network performances. However, most current researches on self-similar traffic mainly focus on the estimation of Hurst index and its impact on network performances, which is not overall. In this paper, the key parameters impacting the network performances of self-similar traffic are investigated. The impact of Hurst index and variance coefficient on network performance is studied by mean of simulation. Analytical results demonstrate that both Hurst index and variance coefficient have great impact on performances. The reason for the impact of variance on performances is analyzed. The relationship and its calculation of $c_\gamma$ and variance are studied. The estimation algorithm of parameters in Superposition of Fractal Renew Process (Sup_FRP) based on Index of Dispersion for Counts (IDC) is proposed. Finally, the impact of fractal onset time on performances is analyzed.

**Keywords**　Network traffic, Self-similarity, Fractional brownian motion, Superposition of fractal renew process (Sup_FRP), Index of dispersion for counts (IDC)

## 1　引言

自 1993 年 Lelan 等人[1]发现网络业务的自相似和长相关特性以来，研究人员进行了大量的有关自相似流量建模和排队性能的分析[2]。研究结果表明，自相似输入下的队列长度分布呈双曲衰减(hyperbolic decay)，而非传统 Markov 模型的指数衰减。自相似和长相关的普遍存在对传统的基于 Poisson 或 Markov 的流量建模、性能分析和流量工程提出了挑战，需要对目前基于 Poisson 或 Markov 的流量工程技术进行重新检验。虽然所有的研究人员都承认网络流量普遍存在自相似和长相关特性，但有关自相似和长相关特性对网络性能到底有没有影响？影响到底有多大？如何影响等问题却有不同的观点。大多数文献认为自相似和长相关特性对网络性能具有重要的影响，需要引入新的业务模型[1~4]；而有的文献则认为实际网络流量虽然确实存在自相似和长相关特性，但从工程实际的角度来看，完全可以用传统的短相关业务模型(如 AR(1)、MMPP 等)来建模[6~7]。

另外，因为 Hurst 系数表示自相似的存在及程度，目前的自相似业务建模和性能评价方法大都片面强调流量过程受 Hurst 系数的影响，而忽略其它因素，这是很不合理的，事实上，虽然 Hurst 系数对网络性能有重要影响，但单靠 Hurst 系数并不能全面反映网络的性能，“Hurst 系数越大，自相似程度越高，网络性能越差”这种大家熟知的结论是有条件的，并且 Hurst 系数本身对网络性能的影响还受其它因素(如流量的方差、缓冲区长度、利用率等)的制约。

在自相似业务识别与参数估计方面，目前的研究基本上全部集中在 Hurst 系数的估计上，而不考虑其它参数。事实上，流量的方差、分形开始时间(fractal onset time)等参数对网络的性能也存在重大的影响，因此还需要估计这些参数。

针对上述存在问题，本文深入研究影响网络性能的自相似流量关键特性，通过 MATLAB 和 OPNET 相结合的仿真方法研究 Hurst 系数和方差系数对网络性能的影响，结果发现 Hurst 系数和方差系数对网络性能均有重要的影响。然后基于长相关的定义，分析方差对网络性能影响的原因，研究 $C_\gamma$ 与方差之间的关系及其计算方法。最后，针对另外一个目前常用的复合分形更新过程模型(superposition of fractal re-

*)西南交通大学科学研究基金项目(2005A03)；国家自然科学资金项目(No. 90104002)。**谭献海**　副教授，博士生，主要研究方向为计算机网络；**黎燕敏**　硕士研究生，主要研究方向为计算机网络；**潘启敬**　教授，主要研究方向为计算机网络；**金炜东**　教授，博导，主要研究方向为职能信息处理、系统仿真。

new process：Sup-FRP)，给出了基于 IDC 的复合分形更新过程参数的估计算法，分析了分形开始时间对网络性能的影响。

## 2 基于 FBM 的自相似流量性能分析

目前有关自相似流量对网络性能影响的研究主要集中在 Hurst 系数的影响上，而忽略其它因素的影响，这是不全面的。事实上，Hurst 系数对网络性能有重要影响，但单靠 Hurst 系数并不能全面反映网络的性能，并且 Hurst 系数本身对网络性能的影响还受其它因素(如流量的方差、缓冲区长度等)的制约。下面通过采用 MATLAB 和 OPNET 相结合的仿真研究影响网络性能的自相似流量关键因素。仿真时参照 pAug. TL 实测流量序列[12]的统计参数：$H=0.873$，平均到达率 $m=2279$ kbit/s，方差系数 $a=262.8$ kbit. s。

图 1 所示的是服务速率 $C=4800$ kbit/s，即利用率 $m/C=0.47$ 时，时延与缓冲区长度之间的关系曲线。由图可以看出，缓冲区长度较小时，时延随缓冲区长度线性增加，然后逐渐稳定在一个固定的值上。但当缓冲区较小时，时延与 Hurst 系数呈反比关系，即 Hurst 系数越大，时延越小，这与一般的理解刚好相反。这是因为有限缓冲区具有重置效应(resetting effect)和截断效应(truncating effect)。重置效应是指当缓冲区为零或清空时，后来的报文将与前面的报文不相关，从而削弱 LRD 的影响。只有当缓冲区中的报文比较多时，长相关才起作用，LRD 产生的影响才明显。类似地，有限缓冲区的截断效应指的是当缓冲区满以后，后来的报文将会被丢弃，从而也会削弱 LRD 的影响。所以，当缓冲区较小时，网络性能将由短相关特性支配。反之，当缓冲区较大时，性能由长相关支配，此时 Hurst 系数越大，时延将越大。为了验证我们的分析结果，增大缓冲区长度，发现当缓冲区长度增大到 6000kb 左右时，Hurst 系数大于 0.9 的平均时延曲线开始发生状态变化，如图 2 所示。即当缓冲区小于 6000kb 时，Hurst 越大，时延越小；而当缓冲区大于 6000kb 时，Hurst 越大，时延越大。Hurst 系数小于 0.9 时也会发生类似的状态转换，但状态转换发生在更大的缓冲区处。

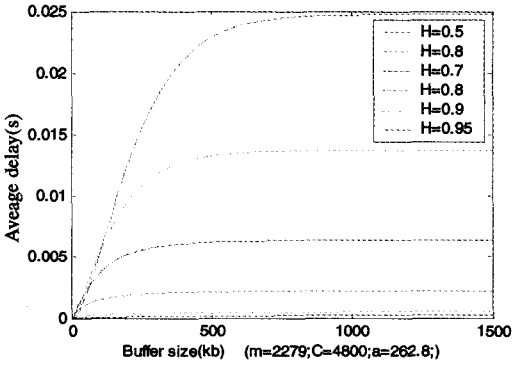


图 1 小缓冲区时平均时延与缓存之间的关系

从图 1 和 2 可以看出，单靠 Hurst 系数并不能全面反映自相似流量对网络性能的影响。而事实上，根据我们的仿真分析，流量的方差对网络性能的影响比 Hurst 系数还要大。

图 3 显示的是平均时延与方差系数之间的关系，其中 Hurst 系数固定为 0.78。由图可以看出，方差对时延的影响受缓冲区长度的制约：缓冲区长度较大时，随着方差系数的增大，平均时延基本上是指数增大，然后稳定在某一个值上。可见缓冲区较大时，方差系数对网络性能具有重要的影响。反之，当缓冲区较小时，随着方差系数的增加，刚开始时延有所增加，但很快就稳定在一个较小的时延值上，即缓冲区较小时，方差系数对时延的影响不大。这是因为较小的缓冲区具有重置效应(resetting effect)和截断效应(truncating effect)，不能记忆流量的变化。

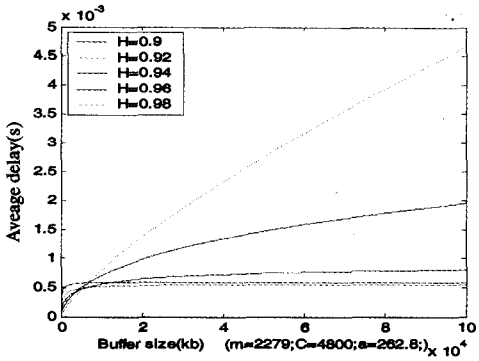


图 2 大缓冲区时平均时延与缓存之间的关系

我们还研究了 Hurst 系数和方差系数对网络性能的不同影响程度，图 4 是 Hurst 系数和方差系数对丢包率影响程度的研究结果，由图可见，方差系数对丢包率的影响程度要比 Hurst 系数大得多。

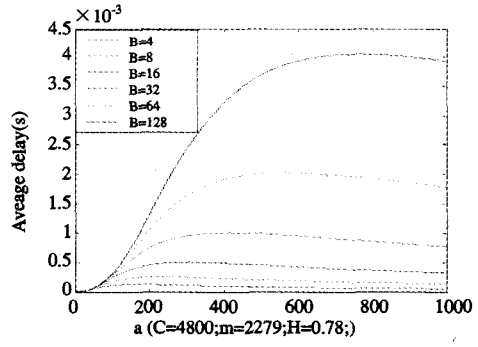


图 3 平均时延与方差系数之间的关系

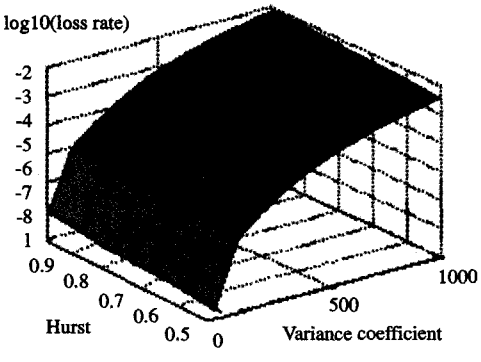


图 4 Hurst 系数和方差系数对丢包率的影响

## 3 影响网络性能的自相似流量关键因素分析

在分形布朗运动(FBM)模型的三个参数$\{H,m,a\}$中，流量平均速率 $m$ 描述流量的“量”特性，而 Hurst 系数 $H$ 和方差系数 $a$ 描述流量的“质”特性，自相似流量对网络性能的影响主要通过 $H$ 和 $a$ 来体现，两者同样重要，这可以通过长相关的一般定义看出。对于一个平稳的二阶自相似过程，其自协方差函数为

$$r(k)\sim c_\gamma |k|^{2H-2}, k\to\infty \quad (1)$$

即自协方差函数除了与 Hurst 系数有关外，还与系数 $c_\gamma$ 有关。如果 $c_\gamma$ 很小，不管 Hurst 系数多大，自协方差函数值都很小。所以，目前有关自相似业务对网络性能的影响主要集中在 Hurst 系数的影响上是不全面的。

根据文[9]的结论，对于长相关(LRD)过程，传统的样本方差渐近表达式 $\sigma_x^2/n$ 变为 $(2c_\gamma n^\alpha/(1+\alpha)\alpha)\cdot(1/n)$，其中 $n$ 为样本大小，$H=(1+\alpha)/2$。可见样本方差与 $c_\gamma$ 成正比关系。

方差对长相关过程的性能有重要影响。根据文[4,5]的研究结果，对于一个平稳 LRD 输入的存储模型，其队列长度 $V$ 的尾分布为 Weibullian 型重尾分布：

$$P(V>x)=O(\exp(-k^2\lambda^{2H}x^{2(1-H)}/2\sum^2)) \tag{2}$$

其中，$\lambda$ 是一个与系统的利用率相关的参数；$k=H^{-H}(1-H)^{H-1}$，$H$ 为自相似系数；$\sum^2=c_\gamma/(H(2H-1))$。可见，对于相对固定的 Hurst 系数，$c_\gamma$ 增大时，队列长度的尾分布增大，从而增大排队时延，这与上一节的仿真结果相符。另外，[0，$t$]时间内到达的负载 $W(t)$ 的方差为 $Var(W(t))\sim\sum^2 t^{2H}$。可见 $c_\gamma$ 与流量样本的方差有关，但不完全相等。

## 4 $c_\gamma$ 与方差系数之间的关系

对于 FBM 模型[3]：

$$A_t=mt+\sqrt{am}Z_t,t\in(-\infty,\infty) \tag{3}$$

其自协方差函数为

$$r(k)=\frac{am}{2}(|k+1|^{2H}-2|k|^{2H}+|k-1|^{2H}),k\geqslant 1 \tag{4}$$

其中 $H$ 为 Hurst 参数，$H\in(0.5,1)$。

将 $(1\pm1/k)^{2H}$ 在二阶上做 Taylor 展开，并取前面三项，得

$$(1\pm\frac{1}{K})^{2H}=1\pm\frac{2H}{k}+\frac{2H(2H-1)}{2k^2} \tag{5}$$

由式(4)、(5)得

$$r(k)\approx amH(2H-1)k^{2H-2} \tag{6}$$

比较式(1)和(6)，得

$$c_\gamma=amH(2H-1) \tag{7}$$

对于长度为 $N$ 的样本，其均值为

$$m=\frac{1}{N}\sum_{i=1}^{N}x_i \tag{8}$$

样本方差 $\sigma_X^2$ 为

$$\sigma_X^2=\frac{1}{N-1}\left[\sum_{i=1}^{N}x_i^2-\frac{1}{N}(\sum_{i=1}^{N}x_i)^2\right] \tag{9}$$

方差系数为

$$a=\frac{\sigma_X}{m} \tag{10}$$

## 5 基于 IDC 的自相似流量关键参数估计算法

在自相似流量模型中，除了分形布朗运动模型外，另一个常用的模型是分形更新过程，其中最常用的是由 Bo Ryu[8]提出的复合分形更新过程模型(superposition of fractal renew process：Sup-FRP)，它可简单直观地刻画网络中出现的自相似现象。该模型仅需三个参数即可描述长相关过程 LRD 的突发业务特性，这些参数具有很强的物理意义。

在复合分形更新过程模型中，三个参数分别为分组平均到达速率 $m$、Hurst 参数 $H$ 和分形开始时间 $T_0$。$m$ 定义了到达过程的平均速率(分组/秒)。$H$ 定义了自相似业务流的自相似和长相关特性，它是自相似性程度的度量，用于描述复合随机过程的突发性。$T_0$ 定义了一个时间尺度，自相似(分形)现象在这个时间尺度上开始出现。$T_0$ 越小，流量的突发程度越高。

复合分形更新过程(Sup-FRP)是由 $M$ 个广义平稳的分形更新点过程(FRP)叠加而成，每个分形更新过程中的事件到达时间间隔服从相同重尾分布：

$$p(t)=\begin{cases}rA^{-1}e^{-t/A}, & 0\leqslant t\leqslant A\\ re^{-r}A^r t^{-(r+1)}, & t>A\end{cases} \tag{11}$$

式中 $1<r<2$ 称为形状参数，用来确定该过程的均值和方差。$A$ 称为位置参数，用来确定该过程可以取的最小值。

实际应用中，确定一个自相似业务源所需参数是平均包到达速率 $m$、自相似参数 $H$ 和分形开始时间 $T_0$，这三个参数与式(11)中三个参数($r$，$A$，$M$)具有如下关系[8]：

$$H=(3-r)/2 \tag{12}$$

$$m=Mr[1+(r-1)^{-1}e^{-r}]^{-1}/A \tag{13}$$

$$T_0^\alpha=2^{-1}r^{-2}e^{-r}(r-1)^{-1}(2-r)(3-r)[1+(r-1)e^r]^2A^\alpha \tag{14}$$

其中，$\alpha=2-r$。

常用的估计 Sup-FRP 参数的算法有功率谱密 PSD (Power Spectral Density)、叠合率 CR (Coincidence Rate)、计数离差系数 IDC (Index of Dispersion for Counts)(也称作 Fano factor)和基于计数的方差函数 COV (count-based CO-Variance function)。下面介绍基于 IDC 的 Sup-FRP 参数估计算法。

对于给定的时间间隔 $T$，自相似流量的 IDC 定义为

$$IDC(T)=\mathrm{var}[\sum_{i=1}^{100T}X_i]/E[\sum_{i=1}^{100T}X_i]\sim cT^{2H-1} \tag{15}$$

其中 $c$ 为有限正常数，$X_i$ 表示每 10ms 时间间隔内到达的报文数。对于自相似过程，$\log(IDC(T))$ 与 $\log(T)$ 之间的关系曲线为一条渐近直线[1]，斜率为 $2H-1$。文[11]给出了另外一个更为简洁的 IDC($T$)和 Hurst 系数之间的关系：

$$IDC(T)=1+(T/T_0)^{2H-1} \tag{16}$$

两边取对数，得

$$\log_{10}\{IDC(T)-1\}=(2H-1)\log_{10}(T)-(2H-1)\log_{10}(T_0) \tag{17}$$

令 $T=1$，得

$$\log_{10}\{IDC(T)-1\}=(1-2H)\log_{10}(T_0) \tag{18}$$

其中 $T_0$ 为分形开始时间。

基于 IDC 估计 Sup-FRP 的 Hurst 系数 $H$ 和分形开始时间 $T_0$ 的算法如下：

1)绘制 $\log_{10}\{IDC(T)-1\}$ 与 $\log_{10}\{T\}$ 之间的关系曲线。对于自相似过程，流量的 IDC 将在所有的时间范围内随时间单调递增，而 Poisson 流量的 IDC 将会很快收敛于某一个固定值；

2)$\log_{10}\{IDC(T)-1\}$ 与 $\log_{10}\{T\}$ 之间的关系直线的斜率等于 $2H-1$，据此可以确定 Hurst 系数的值；

3)令 $T=1$，求对应的 $\log_{10}\{IDC(T)-1\}$ 值，根据 $\log_{10}\{IDC(T)-1\}=(1-2H)\log_{10}(T_0)$ 计算 $T_0$ 的值。

图 5 所示的是具有相同的 Hurst 系数和不同的分形开始时间 $T_0$ 的 IDC 图。分形开始时间 $T_0$ 对网络性能具有重要的影响。图 6 所示的是具有相同 Hurst 系数、不同分形开始

(下转第 48 页)

### 3.4 随机数的安全性问题

PGP 使用两个伪随机数发生器(PRNG):一个是 ANSI X9.17 发生器,另一个是从用户击键的时间和序列中计算熵值从而引入随机性。

• 用户击键引入随机性:这是真正的随机数,只是尽量使击键无规则就行。

• ANSI X9.17 PRNG:使用 IDEA 而不是 3DES 来产生随机数种子。X9.17 需要 randseed.bin 中的 24 bytes 的随机数,PGP 把其他 384 bytes 用来存放其他信息。Randseed.bin 文件最初是利用用户击键信息产生的,每次加密前后都会引入新的随机数,而且随机数种子本身也是加密存放的。

• X9.17 用 MD5 进行预洗:所谓"洗"就是指像洗牌一样把数据打乱。加密前叫预洗,加密后为下一次加密的准备,叫后洗。PGP 的日常随机数产生器 X19.7 是用明文的 MD5 值来预洗的,它基于攻击者不知道明文这样一个假设。

• randseed.bin 的后洗操作:后洗操作被认为是更安全的。更多的随机字节被用来重新初始化 randseed.bin 文件,它们被用当前的随机临时 PGP 密匙来加密。同样,如果攻击者知道这个密匙,他就不用攻击 randseed.bin 文件。相反,他更关心 randseed.bin 文件当前的状态,因为可能从中获得下次加密的部分信息。因此,对 randseed.bin 文件的保护和公匙环及私匙环文件同样重要。

## 参 考 文 献


1 王育民,刘建伟,等.通信网的安全—理论和技术[M].西安电子科技大学出版社,2002

2 郑丽娟,刘莉,等.邮件加密软件 PGP 的安全技术研究与应用[J].河北省科学院学报,2005(4)

3 刘雅丽.PGP 保护电子邮件的研究[J].孝感学院学报,2005(3)

4 杨宗德,等.基于 PGP 的安全电子邮件系统设计与实现[J].信息安全与通信保密,2005(9)

5 魏洪波,周建国,等.安全电子邮件协议[J].现代电信科技,2002(3)

6 陈勇.安全电子邮件系统的设计与分析[J].舰船电子工程,2006(4)

7 郑丽娟,郑丽伟,等.邮件加密算法 PGP 的改进[J].河北大学学报(自然科学版),2004(3)


(上接第 30 页)

时间时平均时延随缓冲区长度的变化曲线,可见分形开始时间越小,网络的性能越差。

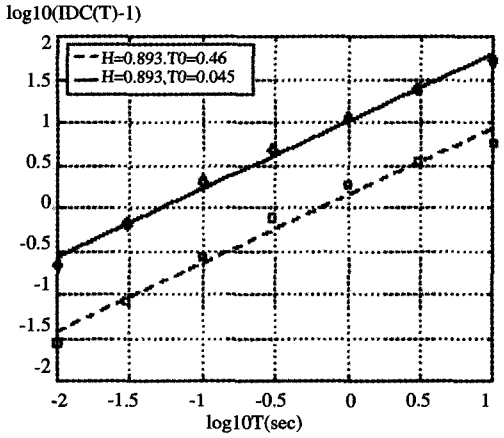


图 5 网络流量的 IDC 图

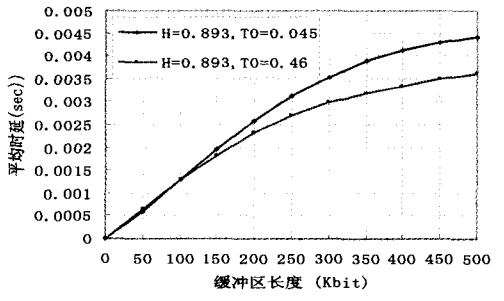


图 6 相同 $H$ 不同 $T_0$ 时的时延

**结束语** 目前的自相似业务建模和性能评价方法大都片面强调流量过程统计特性的某些方面的影响,而忽略其它因素。而事实上,虽然 Hurst 系数对网络性能有重要影响,但单靠 Hurst 系数并不能全面反映网络的性能,同时 Hurst 系数本身对网络性能的影响还受其它因素(如流量的方差、缓冲区长度、利用率等)的制约。同样,在自相似业务识别与参数估计方面,目前的研究基本上全部集中在 Hurst 系数的估计上,而不考虑其它参数。

本文深入研究影响网络性能的自相似流量关键特性,通过 MATLAB 和 OPNET 相结合的仿真方法研究 Hurst 系数和方差系数对网络性能的影响,结果发现 Hurst 系数和方差系数对网络性能均有重要的影响。然后基于长相关的定义,分析方差对网络性能影响的原因,研究 $c_\gamma$ 与方差之间的关系及其计算方法。最后给出了基于 IDC 的复合分形更新过程参数的估计算法,分析了分形开始时间对网络性能的影响。

## 参 考 文 献


1 Leland W E, Willinger W, Taqqu M S, et al. On the self-similar nature of Ethernet traffic (extended version)[J]. IEEE/ACM Trans. Networking, 1994, 2(1): 1~15

2 Park K, Willinger W. Self-similar Network Traffic and Performance Evaluation[M]. New York: Wiley Interscience, 2000

3 Norros I. On the use of fractional brownian motion in the theory of connectionless networks[J]. IEEE J Select Areas Commun, 1995, 13(6): 953~962

4 Norros I. A storage model with self-similar input [J]. Queuing Syst, 1994, 16: 387~396

5 Brichet F, Roberts J, Simonian A, et al. Heavy traffic analysis of a storage model with long-range dependent on/off sources [J]. Queuing System. their Applications, 1996, 23: 197~215

6 Yoshihara T, Kasahara S, Takahashi Y. Practical Time-scale Fitting of Self-similar Traffic with markov-Modulated Poisson Process [J]. Telecomm. Systems, 2001, 17(1-2): 185~211

7 Erramilli A, Narayan O, Neidhardt A, et al. Performance impacts of multi-scaling in wide m a tcp/ip traffic[J]. In: INFOCOM 2000, Tel Aviv, Israel, 2000, 1: 352~359

8 Ryu B K, Lowen S B. Point Process Approaches to the Modeling and Analysis of Self-similar Traffic — Part I: Model Construction [A]. In: Proc IEEE INFOCOM' 96 [C]. San Francisco, March 1996. 1468~1475

9 Beran J. Statistics for Long-memory Processes. London, U K: Chapman & Hall, 1994

10 Watagodakumbura C, Jennings A, Shenoy N. Absolute effects of aggregation of self-similar traffic on quality of service parameters [J]. In: First International Symposium on Control, Communications and Signal Processing, 2004. 511~514

11 Lowen S B, Teich M C. Fractal renewal processes generate 1/f noise [J]. Physics Review E, 1993, 47: 992~1001

12 Lawrence Berkeley National Laboratory. BC-Ethernet traces of LAN and WAN traffic [DB/OL]. http://ita.ee.lbl.gov/html/contrib/BC.html, 2003-06