第 35 卷 第 2 期 计 算 机 工 程 2009 年 1 月
Vol.35 No.2 Computer Engineering January 2009
·网络与通信· 文章编号：1000—3428(2009)02—0083—03 文献标识码：A 中图分类号：TP393

# 长相关网络流量 Hurst 指数估计算法

张 博，汪斌强，智英建
(国家数字交换系统工程技术研究中心，郑州 450002)

**摘 要：**针对传统长相关网络流量 Hurst 指数估计算法估计结果不准确、可变信息受损严重的情况，提出时域内滑窗时变方差之差 Hurst 指数估计算法，采用已知参数的人工分形高斯噪声序列及 Bellcore 采集的真实网络流量序列 BC-pOct89 对其进行验证。结果表明该算法减少了可变信息损失，能动态地刻画全域上的长相关特性，具有较高的准确性和鲁棒性。

**关键词：**长时相关；滑窗时变；分形高斯噪声；鲁棒性

# Long Range Dependent Networks Traffic Hurst Exponent Estimate Arithmetic

ZHANG Bo, WANG Bin-qiang, ZHI Ying-jian
(National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002)

**【Abstract】** Because result of long range dependent networks traffic Hurst exponent estimate arithmetic is not exact and loses much changed information, this paper proposes Slide Window Time Variety(SWTV) variance's dispersion Hurst exponent estimate arithmetic, which uses Fractal Gauss Noise(FGN) list whose parameters is known and real networks traffic list BC-pAug89 collected by Bellcore to test it. The result indicates this arithmetic reduces loss of changed information, depicts Long Range Dependent(LRD) characteristic in all field, and gets well veracity and robust .

**【Key words】** Long Range Dependent(LRD); slide window time variety; Fractal Gauss Noise(FGN); robust

## 1 概述

传统的通信网络业务流模型一般基于Poisson(连续时间)或Bernoulli(离散时间)过程，它们的业务量是短时相关(Short Range Dependence, SRD)的[1]。

国内外学者近年来发现，真实的网络业务在很长时间范围内都具有相关性，即业务流到达为长时相关(Long Range Dependence, LRD)模型，而自相似(self similar)模型是最能够描述这种特性的模型之一。

在自相似网络模型中，Hurst指数能最好地反映这种自相似特性。本文基于以往Hurst指数估计方法，提出了时域内滑窗时变方差之差Hurst指数估计算法，并采用人工分形高斯噪声序列及采集的真实网络流量序列BC-pOct89[2]对该算法的特性进行了验证。

## 2 传统的 Hurst 指数估计方法

### 2.1 时域内 Hurst 指数估计算法

(1)方差法

聚合时间序列$X^{(m)}$的方差具有慢衰减特性，在$m$ 足够大时，满足

$$Var\left(X^{(m)}\right)\propto Var(X)m^{-\beta}$$

其中，Hurst 指数为$H$=1-$\beta$/2。若序列具有长相关性，则$Var(X^{(m)})$与聚合级数$m$在对数图上呈斜率为-$\beta$的直线，且 $\beta\in(-1,0)$。

(2)R/S 估计方法

自相似过程具有 Hurst 效应，即对一给定观测过程$(X_k,k=1,2,\cdots)$，记样本均值为 $\bar{X}_{(n)}$，样本方差为 $S^2(n)$

R/S 方法统计量为

$$R(n)/S(n)=\frac{1}{S(n)}\left(\max(0,w_1,w_2,\cdots,w_n)-\min(0,w_1,w_2,\cdots,w_n)\right)$$

其中，$w_k=(X_1+X_2+\cdots+X_k)-k\bar{X}(n)$，$k$=1,2,$\cdots$,$n$，从长时相关过程观测的数据满足

$$E\left(R(n)/S(n)\right)\sim cn^{H},n\to\infty$$

如果数据是短相关的，则：

$$E\left(R(n)/S(n)\right)\sim dn^{0.5},n\to\infty$$

其中，c, d 为与 $n$ 无关的常数。在对数坐标下，画出 R/S 曲线并进行最小二乘直线拟合可得 Hurst 指数的估计。

(3)小波方法

基于小波变换的 Hurst 指数估计方法可以应用于各个观察尺度。文献[3]提出了对突发业务进行多分辨率采样和小波分解的 Hurst 估计方法，其中，多分辨率采样可减少参数估计所需业务到达过程计数样本总量对采样数据的正交；小波分解可得到不同尺度下小波系数的方差序列。

### 2.2 频域内 Hurst 指数估计算法

(1)Whittle 估计方法

时域估计方法缺乏对业务数据精确的统计，Whittle 的最大似然估计则可解决上述问题，设$X$的谱密度函数$f(x,\boldsymbol{\theta})$=$\delta_\varepsilon^2 f(x,(1,\eta))$，其中，参数向量$\boldsymbol{\theta}$=$(\delta_\varepsilon^2,\eta)$=$(\delta_\varepsilon^2,H,\theta_3,\cdots,\theta_k)$,$\theta_3,\cdots,\theta_k$表征业务的短时相关结构；$\delta_\varepsilon^2$表征描述短时相关AR过程的方

**基金项目：**国家“863”计划基金资助项目(2004AA103130)
**作者简介：**张 博(1982－)，男，硕士，主研方向：通信与信息系统，网络流量模型；汪斌强，教授、博士生导师；智英建，博士研究生
**收稿日期：**2008-07-06 **E-mail：**boz01001 @163.com

差对$\eta$的估计，即使下式最小化：

$$Q(\eta)=\int_{-\pi}^{\pi}\frac{I(x)}{f(x;(1,\eta))}\mathrm{d}x$$

其中，$I(x)$为 $X$ 的周期图。

这种估计方法也可对一段短时业务数据进行分析，但不能检验业务到达过程是否具有长相关特性。

(2)周期图法

对于序列 $\{X(t),t\in R\}$，其周期图为

$$I_N(\lambda)=\frac{1}{2\pi}\left|\sum_{t=1}^{N}X_t\mathrm{e}^{-i\lambda_t}\right|^2$$

其中，$\lambda$ 为频率，$0\le\lambda\le\pi$；$N$ 为样本个数。周期图 $I_N(\lambda)$ 是谱密度的一个渐进无偏估计，可由下式计算：

$$I(\omega_k)=\frac{1}{2\pi N}\left[\left(\sum_{j=1}^{N}X_j\cos(\omega_k j)\right)^2+\left(\sum_{j=1}^{N}X_j\sin(\omega_k j)\right)^2\right]$$

其中，$\omega_k=2\pi k/N, k=1,2,\cdots,N/2$。

将点 $(\omega_k, I(\omega_k))$ 绘于对数坐标系中，得到 Periodogram 图。利用最小二乘法拟合通过这些点的直线。记直线的斜率为 $\gamma$，$0<\gamma<1$。Hurst 指数估计值为

$$H=(1+\gamma)/2$$

通过以上 Hurst 指数估计方法在实际中的应用及图 1 给出的不同方法估计 Hurst 指数的差异，得出以下结论：

(1)以上方法是估计 Hurst 指数比较有效的方法，但并不是一种完全正确的方法。

(2)影响网络流量 LRD 序列 Hurst 指数准确估计的主要因素是序列的非平稳性、序列中的周期成分以及序列的采样白噪声。

(3)以上算法的本质都是以不同的方式对时间序列进行求和平均，这样的 Hurst 指数估计算法会使序列本身的高可变性被平滑，可变信息受损。

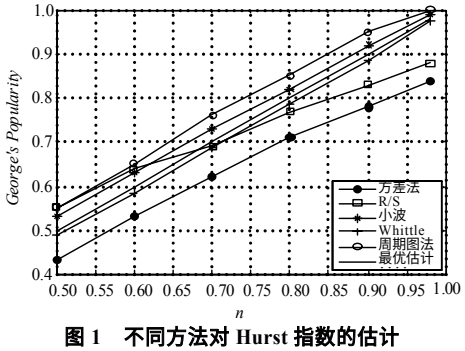


**图 1　不同方法对 Hurst 指数的估计**

本文由此对 Hurst 指数估计算法提出以下建议：(1)用 Hurst 指数进行估值时，必须注明采用哪种估值算法和在何种条件下实施。(2)在判断一序列是否具有 LRD 特性时，应采用多种估值算法，以尽可能提高其判断的准确性。(3)在对序列进行 LRD 特性判定之前，应对序列进行预处理，尽可能消除干扰 LRD 准确判断的成分，如非平稳成分、周期成分、噪声信号。(4)争取发现其他能有效辨识 LRD 特性的参数或者函数。

## 3　滑窗时变方差之差 Hurst 估计算法

本文提出滑窗时变方差之差 Hurst 指数估计算法，并以此为指标对整个序列的 LRD 特性的变化趋势进行定量刻画。

对于网络流量序列 $\{X(k),k=1,2,\cdots,N\}$，定义其局域时移均值序列为

$$\bar{X}_m(k)=\frac{1}{m}\sum_{j=0}^{m-1}X(k+j)$$

其中，$\bar{X}_m(k)$ 是在序列上加了大小为 $m$ 的时窗。

定义 $\{X(k)\}$ 相对于时移均值的样本方差为

$$\sigma_k=\frac{1}{k}\sum_{i=0}^{k}[X(i)-\bar{X}_m(i)]^2$$

方差之差为

$$\sigma_{\Delta i}=\sigma_{k+\delta(i+1)}-\sigma_{k+\delta i}$$

在合适的窗宽下，由不同的 $\bar{X}_m(i)$ 计算的$\sigma_{\Delta i}$与窗宽$m$间具有指数为Hurst的幂指规律，即$\sigma_{\Delta i}\sim cm^{H}$[4]，在对数图上满足线性关系，如图 2 所示。

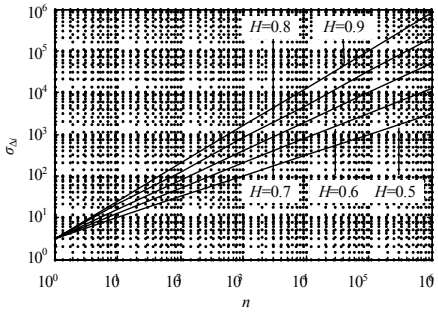


**图 2　$\sigma_{\Delta i}$与$n$的关系**

基于以上定义，大小为$m$的时窗以步径$\delta$在网络流量序列的整个采样域内滑动，每个局域内利用统计量$\sigma_{\Delta i}$计算局域Hurst指数。局域Hurst指数的个数取决于局域宽度$N_{\max}$和窗宽$m$以及滑窗步径$\delta$的大小。

各个局域 Hurst 指数构成整个网络流量序列的全域内时变 Hurst 指数函数，以此为指标定量刻画网络流量的 LRD 特性变化趋势。

该算法具体步骤如下：

(1)预处理。采用短数据域，用$\sigma_{\Delta i}\sim cm^{H}$关系对不同的$m$值和$\delta$值进行估算，在得到的$H$值中密度最大的区域内求均值，与该均值最接近的Hurst值所对应的$m$和$\delta$即为该流量序列的最佳窗宽$m$及合适的滑窗步径$\delta$。

(2)把全时域分成若干个局域，在每个局域内计算时移均值序列 $\bar{X}_m(k)$，再计算$\sigma_k$和$\sigma_{\Delta i}$，用$\sigma_{\Delta i}\sim cm^{H}$通过对数图上直线的斜率计算该局域内的Hurst 指数。

(3)对各个局域上的 Hurst 指数进行多项式插值，形成光滑的曲线，该曲线为该流量序列 Hurst 指数形状的一个逼近。

## 4　性能分析

为了检验该算法对 LRD 特性估计的有效性与鲁棒性，采用该算法分别对已知 Hurst 指数的人工分形高斯噪声 Fractal Gauss Noise(FGN)序列及 Bell- core 采集的真实网络流量数据 BC-pOct89 进行了时变 Hurst 指数的估计。

### 4.1　人工分形序列

本文采用已知 Hurst 指数的人工 FGN 序列，其中，图 3 是该算法对 $H$=0.75 的人工 FGN 序列的估计结果，结果表明，该算法所计算的全时域 Hurst 指数在 0.75 上下浮动，多项式拟合结果与最优估计基本吻合。图 4 为该算法对 Hurst 指数在 0.5~0.9 之间变化的 FGN 序列族的时变 Hurst 指数的估计结果，结果表明，该算法通过多项式拟合后的估计结果与最优估计出现多次重叠。

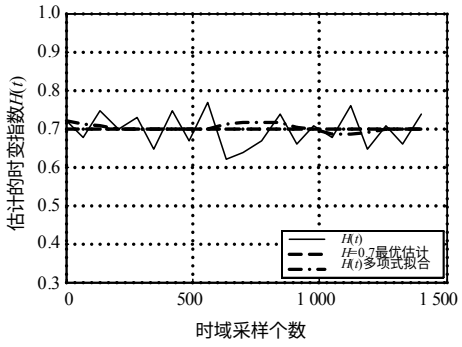


**图 3　*H*=0.75 时人工 FGN 序列估计**

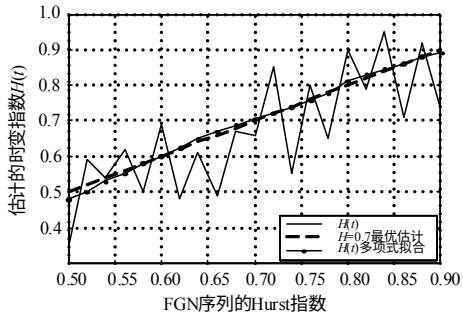


**图 4　Hurst 在 0.5~0.9 变化时 FGN 估计**

在用该算法对 Hurst 指数在 0.5~0.9 之间变化人工序列进行估计时，不同的 Hurst 值应采用不同的窗宽和步长，步长影响不大，窗宽则随着 Hurst 指数的增大不断加宽，方可得到准确的估计结果。

图 1 所示的全域估计方法的估计结果显示，该算法能够动态地刻画全域上的长相关特性。

### 4.2　真实网络流量数据检验

图 5 给出了BC-pOct89 的时域图。BC-pOct89 数据记录了以太网上采集的IP数据包，在用该算法进行分析时，发现窗宽为 28 s能够使$\sigma_{\Delta i} \sim cm^H$关系满足，选择滑窗步径为 15，如图 6 所示，在相同参数下，用该算法对BC-pOct89 数据时变Hurst指数进行估计，多项式拟合的结果表明时变Hurst指数在 0.82 上下波动，这与文献[5]的结果一致。该算法并没有对网络流量序列的性质预先进行特殊的限制，在对全域网络流量进行局域化后，能够克服网络流量的非平稳性[6]对LRD估计的影响，并能体现出信息变化的细节特征。

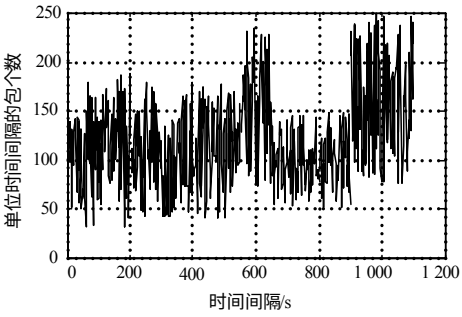


**图 5　BC-pOct89 的时域图**

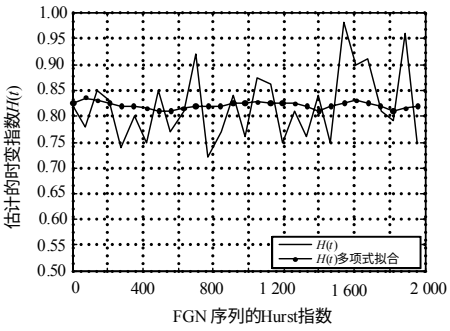


**图 6　BC-pOct89 的 Hurst 指数估计**

### 4.3　结论

本文提出了滑窗时变方差之差 Hurst 指数估计算法的概念，把全域分成若干个局域，按照一定的分辨率实现全域内的 Hurst 指数动态估计，

该算法较以往 Hurst 指数估计算法的优点主要体现在以下几个方面：

(1)不对长相关网络流量序列的性质预先进行特殊限制。

(2)解决了传统 LRD 估计算法在全域内求和平均造成 LRD 信息的损失问题，在局域内加窗求和平均，较好地描述了流量序列中长相关特性的变化趋势。

(3)该算法的多项式插值结果是对全域内不同局域 Hurst 指数的动态估计，能够体现网络流量序列在时域内不断突发变化的细微特征。

(4)适用于分析具有长相关特性的非平稳序列、周期序列以及受噪声影响较大的序列，具有很好的鲁棒性。

## 5　结束语

本文在前人估计 Hurst 指数的各种方法的基础上，从影响 Hurst 指数估计的信号特征方面考虑，提出了滑窗时变方差之差 Hurst 指数估计算法，通过各种人工序列和真实网络流量数据的检验，得出其较传统 LRD 估计算法具有更好的鲁棒性和有效性。

## 参考文献

[1] 王　成, 刘金刚, 刘汉武. 网络业务自相识建模及其 Hurst 系数估计[J]. 计算机工程, 2006, 32(2): 101-102.

[2] Traces Available in the Internet Traffic Archive[EB/OL]. (2004-05-20). http://ita.ee.lbl.gov/html/contrib/BC.html.

[3] 陈惠民, 蔡　弘, 李衍达. 自相似业务: 基于多分辨率采样和小波分析的 Hurst 系数估计方法[J]. 电子学报, 1998, 26(7): 88-104.

[4] Alessio E, Carbone A, Castelli G, et al. Second-order Moving Average and Scaling of Stochastic Time Series[J]. The European Physical Journal B, 2002, 27(2): 197-200.

[5] Leland W, Taqqu M S, Willinger W, et al. On the Self-similar Nature of Ethernet Traffic(Extended Version)[J]. IEEE/ACM Trans. on Networking, 1994, 2(1): 1-15.

[6] Karagiannis T, Molle M, Faloutros M, et al. A Nonstationary Poisson View of Internet Traffic[C]//Proceedings of IEEE INFOCOM'04. Hong Kong, China: IEEE Press, 2004-09.