

网络业务量的一种稳定多分形模型及其检验

张毅 李乐民

(电子科技大学宽带光纤传输与通信网重点实验室 成都 610054)

摘要 目前网络业务量研究中有自相似和多分形两种主要的先验模型。自相似模型简洁,运用单一的 H 指数来描述业务量特性,但它并不能完全表达业务量的信息;多分形模型更准确,但它运用函数形式的多分形谱来刻画信号,不便于实际分析。本文在无穷可分层叠的理论框架下提出了网络业务量的一种稳定多分形模型。运用小波变换方法,通过多个业务量记录的检验,证实此模型从尺度行为和边缘分布两方面都能准确刻画业务量。作为一般多分形过程的特例,该模型把多分形指数的表达式确定下来,并归结为4个参数的估计。

关键词 多分形, 稳定分布, 突发性, 离散小波变换

中图分类号: TN919.2

文献标识码: A

文章编号: 1009-5896(2006)06-1124-05

A Stable Multifractal Model of Network Traffic and Its Verification

Zhang Yi Li Le-min

(Key Lab of Broadband Optical Fiber Transmission and Communication Networks,
UESTC, Chengdu 610054, China)

Abstract There have been two *a priori* models up till now concerning network traffic: self-similar and multifractal models. The self-similar model is simple enough to characterize network traffic with the single exponent H , however it can not thoroughly describe the contents. The multifractal model, which is more accurate, is not practical because it uses multifractal spectra as functions to represent signals. Therefore, on the basis of the framework of IDC theory, this paper proposes a stable multifractal model of network traffic. It is tested on several traces by using the method of wavelet transforms that the model fits traffic precisely in terms of both scale behaviors and marginal distributions. As a specific case of multifractal processes, this model, by estimations of four parameters, can specify the representation of the multifractal exponent.

Key words Multifractal, Stable distribution, Burstiness, Discrete Wavelet Transform (DWT)

1 引言

过去的十年是网络业务量建模取得丰硕成果的时期。最核心成就在于揭示了业务量的各种尺度(Scaling)行为^[1,2]。这大致可以分为自相似(Self-Similar, SS)、多分形(MultiFractal, MF)和无穷可分层叠(Infinitely Divisible Cascades, IDC)3个主要阶段。在IDC的理论框架下, MF模型可看作其特例, 而SS又可当作MF的特例。因而运用IDC可用更广阔的视野来考察业务量。就我们所了解的情况, IDC模型还没有在网络分析中充分应用。本文则以此理论为基础来考察多个业务量记录。通过实际分析我们得出结论, 可以用特殊的MF模型来刻画文中的业务量。

尽管尺度行为的研究是业务量建模的主流, 而业务量的边缘分布也是刻画其特性的一个重要方面, 它直接决定了业务量的突发性(Burstiness)。实际上, 尺度行为就是层叠核分布特性的体现。以往总是将边缘分布和核分布割裂开来研

究。本文则尝试在IDC模型的框架下建立边缘分布和核分布之间的联系。我们通过实际网络业务量估计证实, 文中所采用业务量记录随机过程的绝对值对数和层叠核都服从一类特定的无穷可分(Infinitely Divisible, ID)分布——稳定(Stable)分布。

为体现以上两个方面的内容, 我们提出一种先验模型——稳定多分形(Stable MultiFractal, SMF)过程。文中检验了12个业务量记录都符合该模型。由于建立了业务量尺度行为和边缘分布的联系, 使得只需估计边缘稳定分布的参数就能够捕捉业务量的各方面特性。

小波变换已在工程计算中得到了广泛的采用, 同样也应用于以上各种模型的分析 and 合成中^[3-7]。其采用的小波方法有两种: 小波变换模极大(Wavelet Transform Modulus Maxima, WTMM)和离散小波变换(Discrete Wavelet Transform, DWT)^[8]。本文选用计算复杂度小的后者来分析业务量。

本文的结构如下: 第2节引入SMF模型; 第3节陈述了模型的小波变换估计方法; 第4节是实际网络业务量数据的分析; 最后为结论。

2004-10-14收到, 2005-09-28改回
国家自然科学基金委员会与香港研究资助局联合科研基金
(60218002)资助课题

2 稳定多分形过程

本节首先说明有关的无穷可分层叠和稳定分布的概念,然后叙述我们提出的 SMF 过程。

2.1 自相似、无穷可分层叠和多分形

设有随机过程 $\{X(t), t \in \square\}$, 通常用其平稳的增量过程 (Increment process) $\Delta_l X(t) \square X(t+l) - X(t)$ 刻画业务量单位时间内到达的分组数。自相似模型则表明存在 $H > 0$ 使得:

$$\Delta_{\lambda l} X(t) \stackrel{d}{=} \lambda^H \Delta_l X(t), \quad 0 < \lambda < 1 \quad (1)$$

实际中随机过程的时间应是有限的, 设 $X(t)$ 的 t 取值 $[0, 1]$ 。那么把时间增量 l 当作相对尺度, 就应该小于最大尺度 1, 即 $0 < l < 1$ 。而关系式中的 λ 可以理解为两个不同尺度间的比率。增量过程可以看作特殊的小波变换, $X(t)$ 在不同尺度下的小波变换系数同样服从式(1)。这时将 $\Delta_l X(t)$ 称作在尺度 l 的波动过程 (Fluctuation process)。上式反映了不同尺度 $\Delta_l X(t)$ 边缘分布的倍数关系, 因而通常取其绝对值对数, 转化为随机变量相加关系的研究。定义 $H_l \square \ln |\Delta_l X(t)|$ 作为我们以后分析的主要对象。设 $m(q, l) \square \mathbb{E}[|\Delta_l X(t)|^q]$, 为 $\Delta_l X(t)$ 的绝对值高阶矩, 也可以看作 H_l 的矩母函数 (moment generating function, mgf), 而在文献中亦称作配分函数 (Partition function)^[1]。由式(1)可得到:

$$m(q, l) = K_q l^{qH} \quad (2)$$

其中 $K_q = m(q, 1)$ 。这就是自相似的幂律 (Power-law) 关系。再设 H_l 的概率密度函数 (probability density function, pdf) 为 $p_l(h)$, 则式(1)的关系也可表示为

$$p_{\lambda l}(h) = p_l(h - h_\lambda) = \delta(h - h_\lambda) * p_l(h) \quad (3)$$

其中 $h_\lambda = H \ln \lambda$ 为 λ 决定的常量。

IDC 的概念首先由 Castaing 引入^[9], Arnéodo 在小波分析的框架下将它重新表述^[10], 而 Veitch 等人尝试将它应用到网络业务量分析中^[11, 12]。此模型认为两个不同尺度 l' 和 l 对应的 pdf 之间并非简单的平移关系, 而满足

$$p_{l'}(h) = G_{l', l}(h) * p_l(h) \quad (4)$$

其中函数 $G_{l', l}(h)$ 称作核 (Kernel) 或传播因子 (Propagator), 运算 $*$ 表示卷积。显然有 $\int G_{l', l}(h) dh = 1$, 可以将 $G_{l', l}(h)$ 当作 pdf。并且可认为它是由一个基本核 $G_0(h)$ 多次自卷生成, 而自卷的次数为尺度函数的差:

$$G_{l', l}(h) = [G_0(h)]^{*n(l') - n(l)} \quad (5)$$

于是核所对应的随机变量服从 ID 分布。我们将基本核 $G_0(h)$ 所对应的 ID 分布称作基本 ID 分布。于是配分函数可以表示为

$$m(q, l) = A_q e^{\rho(q)n(l)} \quad (6)$$

其中 $A_q = m(q, 1)e^{-\rho(q)n(1)}$, $\rho(q)$ 为基本 ID 分布的累积母函数 (cumulative generating function, cgf)。当 $n(l) = -\ln l$ 时, 上式简化为多分形的尺度不变 (Scale-invariant) 关系

$$m(q, l) = A_q l^{-\rho(q)} \quad (7)$$

$X(t)$ 就称作多分形过程, $-\rho(q)$ 即为多分形指数 (Multifractal exponent)。若 $-\rho(q)$ 为 q 的线性函数, $X(t)$ 退化为单分形 (Monofractal) 过程。更为特殊的, 若 $-\rho(q) = qH$, 上式进一步简化为自相似过程, 参见式(2)。

2.2 稳定分布

ID 分布是一类广泛的分布, 稳定 (Stable) 分布是其特例^[13]。后面会验证, 我们所设定的随机量 H_l 就可以用稳定分布来估计。用 X, X_1, X_2, \dots 表示一组 i.i.d. 随机变量, 其分布都为 F 。如果对任何 n , 存在常量 $c_n > 0$ 和 $d_n \in \square$, 满足

$$X_1 + \dots + X_n = c_n X + d_n \quad (8)$$

并且 F 不汇聚于一点, 则称 F 为稳定分布。若特征函数表示为 $\mathbb{E}[e^{iuX}] = e^{\psi(u)}$, 则稳定分布的特征指数 $\psi(u)$ 为

$$\psi(u) = \begin{cases} -\gamma^\alpha |u|^\alpha \left[1 - i\beta \tan\left(\frac{\pi\alpha}{2}\right) \text{sgn}(u) \right] + i\delta u, & \alpha \neq 1 \\ -\gamma |u| \left[1 - i\beta \frac{2}{\pi} \ln |u| \text{sgn}(u) \right] + i\delta u, & \alpha = 1 \end{cases} \quad (9)$$

可见, 稳定分布由 4 个参数描述: 稳定指数 (Index of stability) $\alpha \in (0, 2]$, 斜度 (Skewness) $\beta \in [-1, 1]$, 尺度 (Scale) $\gamma > 0$ 和位置 (Location) $\delta \in \square$ 。通常记为 $S_\alpha(\gamma, \beta, \delta)$ 。当 $\alpha = 2, \beta = 0$ 时, 稳定分布退化为正态分布, 此时 γ 和 δ 分别为正态分布的标准差 σ 和均值 μ 。

由稳定分布的特征指数可以得到其累积母函数 $\rho(q) = \psi(-iq)$ 。我们在后面的实际估计中只取 $q > 0$, 且总有 $\alpha > 1.5$, 于是有

$$\rho(q) = \gamma^\alpha \left[1 - \beta \tan\left(\frac{\pi\alpha}{2}\right) \right] q^\alpha + \delta q \quad (10)$$

令 $H_l = \ln |\Delta_l X(t)|$, 有

$$\xi(q) = \gamma^\alpha \left[1 - \beta \tan\left(\frac{\pi\alpha}{2}\right) \right] (q^\alpha - q) \quad (11)$$

显然, $\xi(q)$ 的表达式中比 $\rho(q)$ 少了参数 δ , 在实际检验时也就更简单些。并且, 由后面实际业务量估计可以看到, 式中 $\gamma^\alpha [1 - \beta \tan(\pi\alpha/2)]$ 远小于 δ 。因而 $\rho(q)$ 近似为 q 的斜线, 就不能显示 α 的存在。而 $\xi(q)$ 消去了 δ 的影响, 使其在 $q > 0$ 上为明显的弧线, 突现了 α 的作用。

2.3 稳定多分形过程

显然, IDC 模型关心尺度之间的关系——核 $G_{l', l}(h)$ 为 ID 分布。可由实际的网络业务量计算 $m(q, l)$, 也就可以用式(6)直接估计核分布的参量函数 $\rho(q)$, $n(l)$ 和 A_q 。设边缘分布对应的随机变量 H_l 也为 ID 分布, 且其 mgf 可设为 $m(q, l) = e^{\phi(q)\eta(l)}$ 。代入式(6)可取: $A_q = 1$, $\phi(q) = \rho(q)$, $\eta(l) = n(l)$ 。这样, H_l 与核有相同的由 $\rho(q)$ 确定的基本 ID 分布和尺度函数 $n(l)$ 。至此, 我们建立了边缘分布与核分布/尺度行为之间的联系。只要检测到 H_l 服从 ID 分布, 也就获得了核 $G_{l', l}(h)$ 的信息。

进一步设 H_l 服从稳定分布, 且令尺度行为方面

$n(l) = -\ln l$ ，即多分形形式。我们称此模型为稳定多分形过程。如上述， $A_q = 1$ 。再将式(10)特殊稳定分布的cgf代入式(7)，两边取对数得到：

$$\ln m(q, l) = \gamma_l^\alpha \left[1 - \beta \tan\left(\frac{\pi\alpha}{2}\right) \right] q^\alpha + \delta_l q \quad (12)$$

其中

$$\left. \begin{aligned} \gamma_l^\alpha &= -\gamma^\alpha \ln l \\ \delta_l &= -\delta \ln l \end{aligned} \right\} \quad (13)$$

定义 $\Delta_l X(t)$ 的归一化高阶绝对矩：

$$\begin{aligned} M(q, l) &= \mathbb{E} \left[\left\{ \frac{|\Delta_l X(t)|}{\mathbb{E}[|\Delta_l X(t)|]} \right\}^q \right] \\ &= \frac{m(q, l)}{[m(1, l)]^q} = \frac{A_q}{A_1^q} e^{\xi(q)n(l)} \end{aligned} \quad (14)$$

其中 $\xi(q) = \rho(q) - q\rho(1)$ 。上一节稳定分布的讨论中已经说明 $\xi(q)$ 比 $\rho(q)$ 形式更简化且适合验证。本文将用 $M(q, l)$ 代替 $m(q, l)$ 来进行模型的估计和验证。代入我们 SMF 模型所设的 $A_q = 1$ 及 $\xi(q)$ 和 $n(l)$ 的形式，两边取对数得到：

$$\ln M(q, l) = \gamma_l^\alpha \left[1 - \beta \tan\left(\frac{\pi\alpha}{2}\right) \right] (q^\alpha - q) \quad (15)$$

综上所述，本文将多分形指数的负数为式(10)的多分形过程(参考式(7))称作稳定多分形(SMF)过程。

3 模型的小波变换估计

从统计的角度来看，小波变换方法有着许多优点。譬如：有平稳增量的非平稳过程在小波域上同样能获得平稳性，从而能够采用时间平均来进行估计；通过选择适当的小波消失矩，可以大大减弱随机过程在时间域中的强相关性，以便获得小的估计偏差。本文采用了文献[8]的离散小波变换定义，其小波系数为

$$\begin{aligned} d_j[n] &= \langle f, \psi_{j,n} \rangle \\ &= \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{2^j}} \psi^* \left(\frac{t - 2^j n}{2^j} \right) dt, \quad (j, n) \in \mathbb{Z}^2 \end{aligned} \quad (16)$$

其中

$$\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi \left(\frac{t - 2^j n}{2^j} \right), \quad (j, n) \in \mathbb{Z}^2 \quad (17)$$

为(母)小波函数 $\psi(t)$ 在 $L^2(\square)$ 上构造的一族标准正交基。DWT 的分解和重构有快速“金字塔(Pyramid)”算法。若设离散序列 $f[n]$ 的长度为 N ，则分解和重构的复杂度都为 $O(N)$ 。实际应用中取母小波 $\psi(t)$ 为 9 阶 Daubechies 小波。

显然，实际中业务量序列是有限的，也是离散的。2.1 节已设 $X(t)$ 长度为 1。若实际业务量序列长度为 N ，则可以设其为 $X(t)$ 的以 N^{-1} 为间隙的采样，用 $X[n] = X(N^{-1}t)$ 表示。这样尺度 l 应该满足 $N^{-1} < l < 1$ 。而在实际计算中，通

常将采样间隙标准化为 1。那么进行小波变换的函数为 $f[n] = f(n) = X[n] = X(N^{-1}n)$ ，得到的小波变换系数为

$$\Delta_l X(t) = N^{-1} W f(Nt, Nl) \quad (18)$$

实际计算中尺度通常取二进序列 $\{2^j\}_{j \in \mathbb{Z}}$ ，即 $Nl = 2^j$ 。因此在后面的讨论中应注意，我们所说的尺度 l 是针对原业务量过程 $X(t)$ 的，而级数 j 针对实际用于计算的函数 $f(t)$ 。

4 实际业务量数据分析

4.1 网络业务量记录及其处理

本文所分析的网络业务量数据都取自 ITA(Internet Traffic Archive)^[14]。我们选用了其前 5 组共 12 个基于分组的业务量记录：

BC 4 个从 Ethernet 观察的 LAN 和 WAN 业务量记录；

DEC-PKT 4 个广域网中 1 小时内分组到达记录；

LBL-PKT 2 个广域网中 1 小时分组到达记录；

LBL-CONN-7 1 个广域网中 30 天内 TCP 连接到达记录；

LBL-TCP-3 1 个广域网内 2 小时 TCP 分组到达的记录。

这些记录都包含分组到达时刻。由此我们能够得到分组计数序列，即单位时隙内分组到达的数目。首先针对各业务量选择适当分组计数的最小单位时隙，保证生成的计数过程非零值居多。表 1 中给出了各记录的单位时隙，及其由此得到的分组计数序列长度。参考 2.1 节，这些分组计数序列对应于增量过程 $\Delta_l X(t)$ 。那么进行业务量小波变换的函数应该是原过程 $X(t)$ 。因此在得到这些计数序列后再计算它们的累加过程。

表 1 所测业务量记录

Tab. 1 Traffic traces of the measured networks

记录名称	时隙	长度	α 均值
BC-Oct89Ext	1	122797	1.7494
BC-Oct89Ext4	1	75943	1.7664
BC-pAug89	0.01	314282	1.5590
BC-pOct89	0.01	175960	1.5548
DEC-1	0.01	359994	1.5812
DEC-2	0.01	359980	1.5685
DEC-3	0.01	359979	1.5368
DEC-4	0.01	359978	1.5513
LBL-4	0.01	359999	1.6114
LBL-5	0.01	359999	1.6151
LBL-CONN-7	10	259198	1.6539
LBL-TCP-3	0.05	143999	1.5995

4.2 实际业务量分析

如 2.3 节所述，我们认为 $H_l = \ln |\Delta_l X(t)|$ 服从稳定分布，其归一化矩母函数 $M(q, l)$ 如式(16)。以下用实际网络业务量来验证。首先对上节得到的计数序列的累加过程进行 DWT。我们这里所作的小波变换尺度为二进序列 $\{2^j, j = 1, \dots, 8\}$ 共

8 级。将获得的小波系数 $d_{j[n]}$ 作为波动过程 $\Delta_j X(t)$, 检验其绝对值对数是否服从稳定分布。图 1 以业务量记录 DEC-1 第 1 级 DWT 小波系数为例, 给出了其绝对值对数 pdf, 拟合稳定分布 pdf 和拟合正态分布 pdf 的比较。其拟合稳定分布的 4 个参数分别为 $\alpha = 1.56$, $\beta = -1$, $\gamma_l = 0.58$, $\delta_l = 0.17$ 。而拟合正态分布的 2 个参数分别为 $\sigma = 0.70$, $\mu = 0.08$ 。可以看出有以下两个特点: (1) 波动过程 $\Delta_j X(t)$ 的绝对值对数非常符合稳定分布, 相比之下与正态分布差别较大。因此用对数正态分布来刻画业务量的波动过程是不准确的。(2) 业务量各尺度的斜度参数 β 都近似为 -1 , 这说明波动过程绝对值对数的分布是左半负向的。而非对称的正态分布, 或 $S\alpha S$ 分布。这与业务量单位时间内分组数目小于一个最大值, 且可能为 0 是一致的。

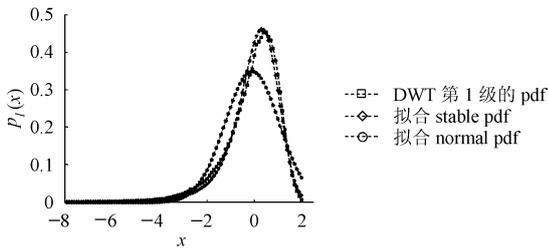


图 1 DEC-1 第 1 级 H_1 的 pdf 比较图

Fig.1 Comparison of pdfs of DEC-1's first level H_1

我们验证了其它 11 个业务量记录也有同样的特征。只是记录 BC-Oct89Ext4(表 2)有些特殊: 它仍然服从稳定分布, 但估计的各尺度 β 都不为 -1 且相差较大, 其原因有待进一步研究。我们猜测是流量异常所导致(如时常断网引起的业务量不平稳, 或有入侵等)。另外, 各业务量作 DWT 估计的各尺度 α 参数有较好的稳定性, 求其均值作为该业务量模型整体的 α 参数值。见表 1 的最后一列, 我们发现如下特点: (1) 稳定指数 α 实际反映了业务量波动过程的突发性: 值越小, 突发性越强。当 $\alpha = 2$ 时, 波动过程为对数正态分布, 突发性最小; (2) 我们所估计的 α 参数与业务量的分组是一致的: 相同网络结构, 记录自不同时间的业务量有相近 α 参数; 而不同组的业务量是有差别的。这说明稳定指数是网络结构的反映; (3) 特别是记录 BC-Oct89Ext4, 前面提到它的各尺度 β 参数值不近似 -1 , 而且彼此差别较大。而从表 1 看到它与记录 BC-Oct89Ext 的 α 参数值相近, 与其它记录有差别。这两个记录采集自同一个网络结构, 只是采集时间不同。因而加强了我们对 α 参数反映网络结构的推测。

应注意到, 我们所估计的是各尺度稳定分布 cgf $\rho_l(q)$, 而并非 $\rho(q)$:

$$\rho_l(q) = \rho(q)n(l) = \gamma_l^\alpha \left[1 + \tan\left(\frac{\pi\alpha}{2}\right) \right] q^\alpha + \delta_l q \quad (19)$$

这里取 $n(l) = -\ln l = (\ln 2)j$, γ_l^α 和 δ_l 应该是级数 j 的线性函数。表 2 给出了业务量记录 DEC-1 在 DWT 后, 各尺度 l / 级数 j 的 H_l 稳定分布参数估计。表中数据显示 α 估计值波动小, 而 β 都为 -1 。图 2 显示估计的 γ_l^α 和 δ_l 随 j 的变化:

表 2 各级 H_l 稳定分布参数估计值

Tab.2 Coefficient estimation of stable distribution on all levels of H_l

j	DEC-1				BC-Oct89Ext4			
	α	β	γ_l	δ_l	α	β	γ_l	δ_l
1	1.56	-1	0.58	0.17	1.79	-0.81	0.83	0.02
2	1.58	-1	0.59	0.56	1.78	-0.66	0.82	0.81
3	1.58	-1	0.59	1.35	1.85	-0.70	0.91	1.65
4	1.58	-1	0.59	2.06	1.79	-0.53	0.91	2.60
5	1.58	-1	0.59	2.89	1.79	-0.34	0.95	3.59
6	1.58	-1	0.60	3.82	1.88	-0.14	0.96	4.76
7	1.61	-1	0.64	4.71	1.67	0.08	0.95	5.91
8	1.58	-1	0.60	5.69	1.58	0.18	0.99	7.28

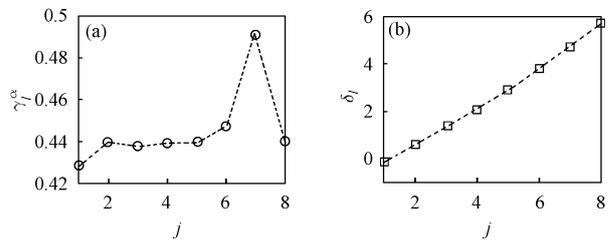


图 2 模型参数 γ_l^α 和 δ_l 随尺度的变化

(a) γ_l^α 随尺度 l / 级数 j 的变化 (b) δ_l 随尺度 l / 级数 j 的变化

Fig.2 Coefficients γ_l^α and δ_l varying with scales

(a) γ_l^α varying with scale l / level j (b) δ_l varying with scale l / level j

δ_l 有明显的线性, 而 γ_l^α 线性较弱。

由稳定分布的估计我们也可以得到:

$$\xi_l(q) = \xi(q)n(l) = \gamma_l^\alpha \left[1 + \tan\left(\frac{\pi\alpha}{2}\right) \right] (q^\alpha - q) \quad (20)$$

由此作出 $\xi_l(q)$ 随 q 的变化图, 见图 3(a)。这里 q 取值以 0.1 为间隔, $0 < q < 2$ 共 21 个值; 其中参数 γ_l 为 DEC-1 第 1 级 DWT 的估计值, 而 α 取其各个尺度估计的平均值。从图 3(b) 可以看到 $\rho_l(q)$ 近似为线性, 即损失了式(19)中项 q^α 的信息。显然不如图 3(a) 中 $\xi_l(q)$ 曲线明显。

以上验证了业务量各级小波变换边缘随机量绝对值对数 H_l 都服从稳定分布, 下面来检验 H_l 与层叠核是否有相同的分布。层叠核的分布由式(14)确定。 $M(q, l)$ 可直接由实际网络业务量估计得到, 作出 $\ln M(q, l)$ 随 q 变化的图, 图 3(a) 显示其与同一尺度 l 下的 $\xi_l(q)$ 基本重合。这印证了我们作出的 H_l 与核同分布的设定。接下来尺度仍然取 8 级二进制序列,

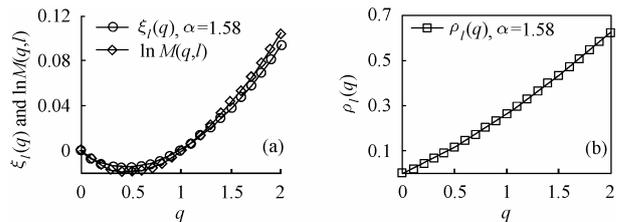


图 3 DEC-1 第 1 级 DWT 的 $\xi_l(q)$ 和 $\rho_l(q)$

(a) $\xi_l(q)$ 随 q 的变化 (b) $\rho_l(q)$ 随 q 的变化

Fig.3 $\xi_l(q)$ and $\rho_l(q)$ of DEC-1's first level DWT

(a) $\xi_j(q)$ varying with q (b) $\rho_j(q)$ varying with q
 q 也取前面的 21 个值, 图 4 显示了 $\ln M(q, l)$ 分别以 q 和 j 为横坐标的变化图。图 4(b) 与 γ_j^α 随 j 变化图一致。由此, 我们验证了 $\xi_j(q)$ 和 $\ln M(q, l)$ 的一致性。

我们对其它的业务量记录用同样的方法进行估计和验证, 都得到同样的结果。

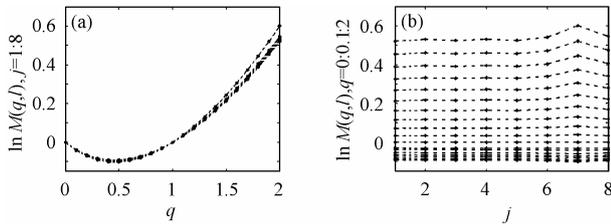


图 4 $\ln M(q, l)$ 随矩阶数 q 和尺度级数 j 的变化
 (a) $\ln M(q, l)$ 随 q 的变化 (b) $\ln M(q, l)$ 随 l/j 的变化
 Fig.4 $\ln M(q, l)$ varying with order q and level j
 (a) $\ln M(q, l)$ varying with q (b) $\ln M(q, l)$ varying with l/j

5 结束语

我们由 IDC 的理论体系提出了 SMF 模型。此模型在尺度行为方面是多分形的, 也就具有了尺度不变性 (Scale-invariance), 这体现在尺度函数 $n(l)$ 可直接表示为 $-\ln l$ 。而其业务量的绝对值对数分布和核分布皆为稳定分布。经小波变换的计算和估计, 我们验证了文中网络业务量都符合该模型。因而由边缘分布估计的稳定分布参数就可确定此模型, 也就刻画了网络业务量的特征: 稳定指数 α 体现了业务量的突发性, 本质上由网络结构所决定; 尺度参数 γ 和平移参数 δ 分别相当于方差和均值, 刻画了业务量的 2 阶和 1 阶特性; 而斜度参数 β 体现了业务量的取值范围, 大多数情况下为 -1 。此模型比自相似模型更准确地刻画了业务量特性, 又将 IDC 模型中两个参量函数的形式解析表示出来, 归结为稳定分布参数的估计。因而为业务量的准确合成和网络性能的理论分析提供了前提条件。

我们注意到已经有基于稳定分布的网络业务量模型^[15-17]。但我们的模型与其有本质的不同: 他们的模型都直接设定过程 $\Delta X(t)$ 的边缘分布是稳定的, 而本模型认为 $H_l = \ln |\Delta X(t)|$ 才是稳定分布的。显然他们的模型是加性而非乘性的, 故不属于 IDC 类型。而 Veitch 等人所讨论 IDC 模型^[11, 12] 中的 $H(q)$ 是我们的稳定分布累积母函数 $\rho(q)$, 但他们用 q 整数次幂的多项式来拟合式(10), 因而不准确。

参考文献

- [1] Abry P, Baraniuk R, Flandrin P, Riedi R, Veitch D. The multiscale nature of network traffic: discovery, analysis, and modelling, *IEEE Signal Processing Magazine*, 2002, 19(3): 28-46.
- [2] Erramilli A, Roughan M, Veitch D, Willinger W. Self-similar traffic and network dynamics. *Proc. IEEE*, 2002, 99(5): 800-819.
- [3] Abry P, Veitch D. Wavelet analysis of long range dependent traffic. *IEEE Trans. on Information Theory*, 1998, 44(1): 2-15.
- [4] Veitch D, Abry P. A wavelet based joint estimator of the parameters of long-range dependence. *IEEE Trans. on Information Theory*, 1999, 45(3): 878-897.
- [5] Riedi R, Crcuse M, Ribeiro V, Baraminuk R. A multifractal wavelet model with application to network traffic *IEEE Trans. on Information Theory*, 1999, 45(3): 992-1018.
- [6] Ma S, Ji C. Modeling heterogeneous network traffic in wavelet domain. *IEEE/ACM Transactions on Networking*, 2001, 9(5): 634-649.
- [7] Tian X, Wu J, Ji C. A unified framework for understanding network traffic using independent wavelet models, in *INFOCOM 2002*, vol. 1. New York, NY, USA, 2002: 446-454.
- [8] Mallat S. *A Wavelet Tour of Signal Processing*, 2nd edition ed. San Diego: Academic Press, 1999.
- [9] Castaing B. The temperature of turbulent flows, *Journal de Physique II*, 1996, 6: 105-114.
- [10] Arneodo A, Muzy J, Roux S. Experimental analysis of self-similar random cascade processes: application to fully developed turbulence. *Journal de Physique II*, 1997, 7: 363-370.
- [11] Veitch D, Abry P, Flandrin P, Chainais P. Infinitely divisible cascade analysis of network traffic data, in *ICASSP 2000*, Istanbul, 2000, 1: 245-248.
- [12] Roux S, Veitch D, Abry P, Huang L, Michell J, Flandrin P. Statistical scaling analysis of TCP/IP data using cascades, in *ICASSP 2001*, vol. 6. Salt Lake City, UT, 2001: 3425-3428.
- [13] Feller W. *An introduction to probability theory and its applications*: John Wiley & Sons, 1971.
- [14] ITA, Internet Traffic Archive, <http://ita.ee.lbl.gov/html/traces.html>.
- [15] Gallardo J, Makrakis D, Orozco-Barbosa L. Use of alpha-stable self-similar stochastic processes for modeling traffic in broadband networks. *Performance Evaluation*, 2000, 40(1): 71-98.
- [16] Karasaridis A, Hatzinakos D. Network heavy traffic modeling using alpha-stable self-similar processes. *IEEE Trans. on Communications*, 2001, 49(7): 1203-1214.
- [17] Laskin N, Lambadaris I, Harmantzis F, Devetsikiotis M. Fractional levy motion and its application to network traffic modeling. *Computer Networks*, 2002, 40(4): 363-375.

张毅: 男, 1974 年生, 博士生, 研究方向为宽带通信网, 网络业务量的建模与分析。

李乐民: 男, 1932 年生, 教授, 博士生导师, 中国工程院院士, 研究方向为数字信息传输与通信网。