

自相似业务量的多重分形分析

张 鹏, 廖建新, 程时端

(北京邮电大学程控交换技术和通信网国家重点实验室, 北京 100876)

摘 要: 本文应用多重分形分析方法对宽带网络中自相似业务量的分形行为进行分析, 与传统的统计测量方法比较, 这种方法可以同时刻划被观察数据序列的短期和长期的分形行为. 通过对一些业务量模型以及真实业务量的分析, 表明该方法可以比较细致地刻划业务量的分形特性, 得到的一个重要的结论是宽带网络中的业务量具有多重分形特性.

关键词: 业务量建模; 自相似性; 分形分析; 统计分析

中图分类号: TN915.01 **文献标识码:** A **文章编号:** 0372-2112 (2000) 01-0096-03

Multifractal Analysis of Self-similar Traffic

ZHANG Peng, LIAO Jian-xin, CHENG Shi-duan

(National Lab. of Switching Technology and Telecommunication Networks,
Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In this paper, we employ an approach of multifractal analysis to characterize the self-similar traffic in broadband networks. In comparison to the traditional statistical analysis, this approach possesses the capability of describing the long-term fractal behavior as well as the short-term behavior. By using this approach, we analyze the real traffic and the results prove the effectiveness of this approach. An important conclusion is obtained that traffic in broadband networks owns the multifractal property.

Key words: traffic modeling; self similarity; fractal analysis; statistical analysis

1 引言

大量的研究表明一些实际网络中的业务量具有长期相关性, 也称为自相似性. 由于分形模型具有自相似性的天然特征, 所以分形模型已成为描述业务量的自相似性的最为重要的模型. 其中最常用也是最为简单的一种分形模型是分形布朗运动, 其业务量的自相似性质由一个 H (Hurst) 参数刻划. 当 H 参数的取值在 $(0.5, 1)$ 之间时, 过程具有自相似性质并且 Hurst 参数值越大, 表明自相似程度越高. 长期相关性对排队性能有着非常重要的影响^[1-3]. 目前, 有关通信网络中业务量分形模型的研究都集中在刻划业务量的长期相关性方面.

然而, 从分形过程的角度考虑, 长期相关性仅仅是分形行为的一个特征, H 参数只能表示出一个分形过程的长期行为. 在真实网络中, 业务量的变化往往是非常复杂的, 对于这样一个复杂的分形过程, 其短期行为可能具有与长期行为不同的分形特征. 就流量控制的应用而言, 了解短时间内业务量的快速突然的变化要比了解业务量的长期相关性质显得更为重要. 另外, 实际业务量并不是一个纯粹的自相似过程, 其自相似性只在一定的时间刻度范围之内有效. 目前, 还没有一个方法可以对业务量的长期和短期行为同时进行分析和刻划, 这影响到人们在观察真实业务量时只能单独考察短期行为和长期行为, 然后建立短期相关模型和长期相关模型, 再将二种模

型合成^[4]. 这种方法不仅增加了建模的复杂程度, 而且还不利于对真实业务量特性的理解.

本文应用多重分形分析方法来同时考察业务量的短期及长期的分形行为, 短期的行为由某一点上的分形指数给出, 而长期行为则由分形指数在整个时间域上的统计分布来刻划. 与传统的测量 H 参数相比, 这一方法可以全面细致地刻划业务量的分形特性. 该方法为研究直接描述长期相关性和短期相关性的业务量的模型打下了基础. 本文第二节对这一方法进行了描述; 在第三节中应用该方法对真实业务量序列进行了分析; 最后对本文进行了总结.

2 多重分形分析方法

在分形理论中, 分形维数用来描述分形集合的复杂性. 多重分形又称作多标度分形或多重分形测度, 对于许多非均匀的分形过程, 一个维数无法描述其全部特征, 需要采用多重分形测度或维数的连续谱来表示. 多重分形维数及其分析方法在物理、地质、材料科学, 特别是在模式识别、自然图像的模拟和信息信号的处理等方面都得到了广泛地应用^[5,6]. 带宽网络中的业务量也是一个复杂的分形过程, 因此应用多重分形维数进行分析将有助于对业务量特性的深入了解. 本节将对多重分形维数及其分析方法进行介绍.

收稿日期: 1998-09-16; 修订日期: 1999-05-24

2.1 多重分形谱和霍尔德(Holder)指数

考虑一个定义在单位区间 $[0,1]$ 上的概率变量 μ 以及它在连续区间 I 上的取样.

$$I_n^k = [k2^{-n}, (k+1)2^{-n}), k=0,1,\dots,2^n-1, n \in \mathbb{N}.$$

由文[7,8]知,如果速率函数

$$\begin{aligned} \tau(q) &:= -1 + \lim_{n \rightarrow \infty} \frac{-1}{n} \log_2 E_n[\mu(I_n^k)^q] \\ &= \lim_{n \rightarrow \infty} \log_2 \sum_{k=1}^{2^n} \mu(I_n^k)^q \end{aligned} \quad (1)$$

存在并且在实数域上可微分,那么有多分形谱函数

$$f_C(\alpha) := \lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 2^n P_n[\alpha(I_n^k) \in (\alpha - \epsilon, \alpha + \epsilon)] \quad (2)$$

存在.其中 $\alpha(I_n^k) := \frac{-1}{n} \log_2 \mu(I_n^k)$,称为 Holder 指数; f_C 描述了 Holder 指数与其期望值的方差.对于一个实序列而言,Holder 指数表示一点(或局部区间)的行为特征,而 f_C 则表示整个区间上的总体行为特征.本文提出的多分形分析方法,通过求 Holder 分形参数和多分形谱函数对业务量序列进行分形分析,从而全面细致地刻划出被考察序列的分形行为.

另外,根据文[9]定义

$$f_L(\alpha) := \inf_{q \in \mathbb{R}} (q\alpha - \tau(q)) \quad (3)$$

有 $f_C(\alpha) = f_L(\alpha) := \tau(\alpha)$. f_L 称为 Legendre(多分形)谱函数. f_L 与 f_C 具有同样的物理意义,但 f_C 式中有二个参变量 ϵ, n ,估计方法比较复杂. f_L 相比之下更容易计算一些.因此,本文提出的多分形分析方法采用 f_L 函数进行序列分析.

为了说明 Holder 指数和多分形谱函数的物理含义,我们考虑一个简单情况.由测度论和分形理论知道,二项式分布满足式(2).考虑 N 重伯努利实验,正数 m_0, m_1 分别表示取值0,1的概率, $m_0 + m_1 = 1$.不失一般性,假定 $m_0 > m_1$.设 N 重伯努利实验的一次实现 $x = (0, \sigma_1, \sigma_2, \dots)_2$, $P[\sigma_k = i] = m_i$.在这种情况下,有 $\mu(I_n^k) = m_{\sigma_1} \cdots m_{\sigma_n}$ 和 $\alpha(I_n^k) = -\frac{1}{n} \cdot \sum_{l=1}^n \log_2 m_{\sigma_l}$.而且当 $n \rightarrow \infty$ 时, $\alpha(I_n^k) \rightarrow \alpha_0 := E[-\log_2 m_{\sigma_1}] = -\frac{1}{2} \log_2 m_0 m_1$, $\alpha_{\min} = -\log_2 m_0$, $\alpha_{\max} = -\log_2 m_1$.根据式(1),可得 $\tau(q) = -\log_2(m_0^q + m_1^q)$.因此产生一个 \cap 形分形谱函数 f_L ,最大值 $f_L(\alpha_0) = 1$,在区间 $(\alpha_{\min}, \alpha_{\max})$ 内是正值.从物理含义上来说,Holder 指数表示某一点 x 上的突发程度,区间 $[x, x + \Delta x]$ 内事件发生的个数近似为 $(\Delta x)^\alpha$.因此,如果 $\alpha < 1$,表示在 x 周围区间内的所有尺度内都有突发,即含有自相似性.而 $\alpha > 1$ 表示随着区间的缩小,事件变得稀疏.有关 Holder 指数的信息都包含在了多分形谱函数中.如果一个过程的多分形谱函数的 $\alpha < 1$ 的区间比较大,那么这个过程就含有较大的多分形成分.

2.2 多重分形谱的估计方法

利用多分形谱函数和 Holder 指数,可以分析序列的分形行为,给出估计多分形谱的方法.考虑一个定义在单位区间 $[0,1]$ 上,解析度为 $N = 2^n$ 的概率变量 μ 的取样,为了检查是否有分形行为存在,我们求不同 q 值和解析度下的部分和

$$S_m(q) = \sum_{k=0, \mu(I_{N/m}^k)^q \neq 0}^{N/m} (I_{N/m}^k)^q, m=1, 2, 2^2, \dots, 2^n$$

如果在 $S_m(q)$ 在 $\log\text{-}\log$ 图上的一段区间内是 m 的线性函数,那么被观察序列在这一区间存在多分形特性,或者说是多分形的.

利用最小二乘区域的方法,可以通过解等式

$$\log S_m(q) \approx \tau(q) \log m + \text{const} \quad (4)$$

求出速率函数 $\tau(q)$.于是由式(3),可以求出多分形谱函数 $f_L(\alpha)$.多分形区间可以通过观察部分和函数增量的线性区间而确定.

2.3 Hurst 参数和多分形分析

统计意义上观察分形行为是基于自相似过程的概念,一个具有平稳增量的自相似过程 $Y_i, Y_{mi} \stackrel{d}{=} m^H Y_i, (0.5 < H < 1)$,

其中符号 $\stackrel{d}{=}$ 表示在概率分布上相同. H 参数用来表示自相似的程度, H 参数值越大,表明自相似程度越高,或者说长期相关性越强.如果是一个严格自相似过程,它在所有时间尺度范围内都保持有相同的 H 值.如果是一个近自相似过程,则在一定尺度范围内具有相同的 H 值.测量 H 参数的方法有很多,在实际应用中为了得到比较准确的测量值,一般是用不同的方法测量 H 值再进行平均.鉴于 H 参数在自相似模型中的重要性,我们考察一种与多分形分析比较相近的测量 H 值的矩统计方法^[10],来说明 H 参数与多分形分析的关系.

设一个具有平稳增量的过程 Y_i ,增量 $X_i = Y_i - Y_{i-1}$.令

$$X_k^m = 1/m \sum_{l=1}^m X_{(k-1)m+l} = 1/m \bar{X}_k^m$$

于是有 $X_k \stackrel{d}{=} m^{1-H} X_k^m$.类似多分形分析

$$S_m^X(q) := \sum_{k=1}^{N/m} |\bar{X}_k^m|^q$$

如果 X 是自相似过程,则有

$$\log S_m(q) \approx \gamma(q) \log m + \text{const}.$$

而且 $\gamma(q) = qH - 1$.于是,可以估计出 H 参数的值.

可见运用类似多分形分析的矩统计方法,可以求出 H 参数.但是矩统计方法求出的 H 值,其所包含业务量的特征是非常有限的.这主要是因为实际业务量的变化是非常复杂的,它的短期和长期的业务量特征都非常丰富.Hurst 参数无法刻划在小时间尺度时业务量所具有的分形特性,而且也无法描述出业务量的变化情况.相比之下,多分形谱函数则包含了丰富的业务量的流量特征,这也正是要引入多分形分析方法的原因.

3 真实业务量的多分形分析

在这一节中,应用多分形方法对真实业务量进行分析.LAN 业务量和 VBR 视频业务量是二种典型的自相似业务量^[11,12].首先分析 LAN 业务量.

Bellcore 的 LAN 业务量是研究自相似模型最常用的数据集,为了使结果有可比性,我们应用数据集 pOCT-TL 和 OCText 数据集进行多分形分析.前者是 LAN 内部业务流量,后者是流出 LAN 的业务流量.定义变量 Z 为单位时间内字节数,最小时间单位为 10 毫秒.首先将数据集进行归一化, $z_k = Z_k / \sum_{i=1}^N Z_i$, N 是序列的长度,然后计算部分和

$$S_m(q) = \sum_{k=1}^{N/m} (\bar{z}_k^q)^q, \text{ 其中 } \bar{z}_k^q(q) = \sum_{l=1}^m z_m(k-1) + l$$

如图 1 绘出序列 pOCT-TL 的 $\log S_m(q) \sim \log m$ 的图形, 从上到下, 是当 $q = -3, -2.5, \dots, 2.5, 3$ 时的曲线。由图中可以看到, 每一条曲线都近似一条直线。这说明该序列有很好的多分析比例特性存在, 分形特性在很大范围内存在。根据公式 (3), 估计 pOCT-TL 的多分形谱函数 $f_L(\alpha)$ 。明显可见, $f_L(\alpha)$ 是一个凸函数。当 $\alpha \approx 1.0$ 时, $f_L(\alpha)$ 取最大值 1。Holder 指数取值在 0.4~1.6 之间时, $f_L(\alpha)$ 为正值。按照第 2.3 节中的概念, 当 $\alpha < 1$ 表示随机过程在所有时间尺度上具有自相似性, 所以 pOCT-TL 序列 (LAN 的内部业务量) 具有较大的多分形成份。文中也给出了序列 OCText 的多分形谱函数。类似地, LAN 的外部业务量也含有较大的多分形成份。另外, 虽然二个序列的多分形谱函数比较相近, 但是仍然有一定的差别, 这可以通过测量特定点的 Holder 指数看出。表 1 中, 列出了参量值 α_0 ,

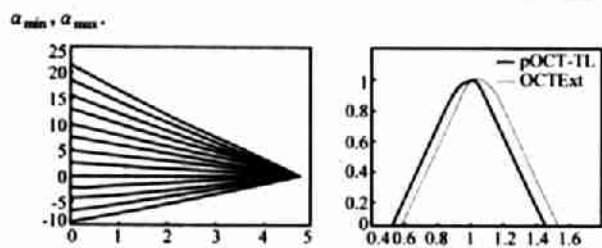


图 1 LAN 业务量 (数据集 pOCT-TL) 的 $\log S_m(q) \sim \log m$

图和多重分形谱函数 $f_L(\alpha)$

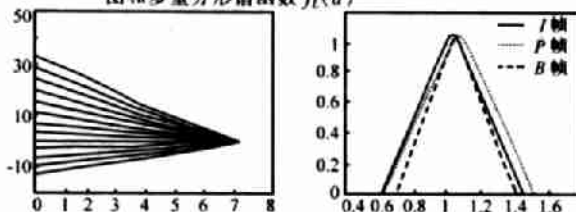


图 2 Starwar (B 帧) 的 $\log S_m(q) \sim \log m$ 图

和多重分形谱函数 $f_L(\alpha)$

下面, 对 VBR 的视频业务量进行分析, 采用 Starwar 的 MPEG 数据集, 对 I 帧、P 帧和 B 帧序列分别求出部分和函数和多分形谱。简洁起见, 这里只绘制 B 帧序列的 $\log S_m(q) \sim \log m$ 图形 (如图 2)。可以发现 MPEG 的业务量的多分形谱函数也呈凸函数形并具有较大的多分形成份。不同帧序列的谱函数曲线不尽相同 (见表 1)。另外, 对照图 1 和图 2, LAN 业务量和 MPEG 业务量的多分形谱函数的形状有明显的差别, LAN 业务量的 $f_L(\alpha)$ 曲线比 MPEG 业务量的曲线形状要宽, Holder 指数小于 1 的成分要大, 也就是 LAN 业务量比 MPEG 业务量所包含的多分形成份要多。通过掌握这些信息, 就可区分业务量的类型。

表 1 真实业务量序列的自相似特征和多分形特征

数据集	序列长度	Hurst 参数	α_0	α_{\min}	α_{\max}
LAN pOCT-TL	2^{16}	0.7857	0.964	0.531	1.430
LAN OCText	2^{16}	0.8332	1.123	0.587	1.547
STARWAR I 帧	2^{12}	0.8618	1.010	0.620	1.522
STARWAR P 帧	2^{14}	0.8746	1.136	0.627	1.463
STARWAR B 帧	2^{16}	0.8946	1.136	0.705	1.401

4 结论

多分形分析方法可以非常细致刻画业务量的分形特性, 它不仅可以描述局部的短期相关性质, 而且还包含了长期相关性质的信息。通过对真实业务量的观察, 我们发现一些实际网络业务量在多个时间尺度上存在很好的多分形特性。而且不同种类的业务量多分形行为不同。因此, 多分形分析是一个很有前途的刻画业务量特性的工具, 有理由相信多分形分析的方法会对业务量的建模和流量控制起到积极地推动作用。本文以后的工作将着重于研究一种基于多分形分析方法的自相似业务量模型。

参考文献

- [1] I. Norros. A Storage Model with self-similar input. *Queueing Systems* 1994, 16: 387 ~ 396
- [2] N. Likharev et al. Analysis of an ATM buffer with self-similar (Fractal) input traffic. *Infocom'95*, 1995
- [3] A. Adas. Traffic models in broadband networks. *IEEE Communication Magazine*, Jul. 1997
- [4] L. M. Kaplan et al. Signal modeling using increments of extended self-similar process. *ICASSP'94*, 1994
- [5] P. R. Jelenkovic, A. A. Lazar, N. Semret. The effect of multiple time scales and subexponentiality in MPEG video streams on queuing behavior. *IEEE JSAC*, Aug. 1997, 15(6): 1052 ~ 1071
- [6] 谢和平, 薛秀谦. 分形应用中的数学基础与方法. 北京: 科学出版社, 1996(1)
- [7] K. J. Falconer. *Fractal Geometry: Mathematical Foundation and Applications*. John Wiley and Sons, New York, 1990
- [8] T. Halsey et al. Fractal measures and their singularities: The characterization of strange sets. *Phys. Rev.* 1986, A22: 1141 ~ 1151
- [9] R. H. Riedi. An improved multifractal formalism and self-similar measures. *J. Math. Anal. Appl.* 1995, 189: 462 ~ 490
- [10] M. Taqqu, V. Teverovsky and W. Willinger. Is the ethernet data self-similar or multifractal? *Fractals* 1997, 5: 63 ~ 73
- [11] Will E. Leland, Murad S. Taqqu, W. Willinger, Daniel V. Wilson. On the self-similar nature of ethernet traffic (Extended Version). *IEEE/ACM Transactions on Networking*, 1994, 2: 1 ~ 15
- [12] Jan Berni, Robert Sherman, Murad S. Taqqu, W. Willinger. Long-range dependence in variable-bit-rate video traffic. *IEEE Transactions on Communications*, 1995, 43: 1566 ~ 1579



张 鹏 1973 年出生, 96 级博士研究生, 研究方向为宽带网络流量控制和业务量建模, 已发表论文十余篇。