

基于分数布朗运动的自相似流量判别及生成方法

张雪媛^{1*}, 王永刚², 张 琮³

(1. 北京工业职业技术学院 信息中心, 北京 100042; 2. 国防大学 信息作战与指挥训练教研部, 北京 100091;

3. 北京航空航天大学 可靠性与系统工程学院, 北京 100191)

(* 通信作者电子邮箱 wangqcl23@163.com)

摘 要: 针对网络流量自相似程度判别方法较少和应用分数布朗运动(FBM)进行自相似流量模拟时可能会产生负值流量等问题, 给出一种基于多阶矩的自相似流量判别方法和改进 FBM 模型的自相似流量模拟方法。首先通过分析样本矩的数学式, 在分形矩分析的基础上得到一种多阶矩的自相似判别方法, 然后对经典的随机中点置位(RMD)算法进行改进, 最后对 Bellcore 和 LBL 实验室采集的真实流量数据进行自相似判别和模拟, 仿真验证实验结果表明该方法的有效性。

关键词: 多阶矩; 随机中点置位算法; 分数布朗运动过程; 自相似性; 判别方法; 生成方法

中图分类号: TN913.23 **文献标志码:** A

Self-similar traffic discrimination and generating methods based on fractal Brown motion

ZHANG Xueyuan^{1*}, WANG Yonggang², ZHANG Qiong³

(1. Information Center, Beijing Polytechnic College, Beijing 100042, China;

2. Department of Information War and Command Training, National Defense University, Beijing 100091, China;

3. School of Reliability and Systems Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China)

Abstract: To deal with the difficulties of lacking the discrimination method of network's traffic self-similarity and producing negative traffic based on classical Fractal Brown Motion (FBM), a discrimination method was proposed based on multiple order moment and a generation method was provided based on modified FBM model. Firstly, the mathematical formula of sample moment was studied. The discrimination method of self-similarity traffic was obtained on account of fractal moment analysis. Secondly, the classical Random Midpoint Displacement (RMD) algorithm was modified. At last, taking account of the real traffic of Bellcore and LBL, the discrimination method and generation method were given. The comparison of the simulation results with the actual experimental data proves that the method is feasible.

Key words: multiple order moment; Random Midpoint Displacement (RMD) algorithm; Fractal Brown Motion (FBM) process; self-similarity; discrimination method; generation method

0 引言

随着互联网快速发展、网络应用多样化以及新型网络应用快速部署, 网络流量特性也随之发生变化。如何抽象出流量的一般化特性成为近期研究的热点。1993 年 Leland 等率先发现网络业务的自相似特性, 此后大量研究结果表明, 自相似、多分形和长相关(Long Range Dependence, LRD)是现代网络业务的普遍特性^[1-3]。网络业务自相似、LRD 特性与基于 Markov 链假设的业务模型有着截然不同的特性, 对网络的性能有着重大的影响, 如果继续使用传统的网络业务模型, 如 Markov 链、Poisson 过程或 Gaussian 过程等来描述已不再适合^[4-7]。

自相似流量牵涉到两个方面的研究, 其一为自相似流量判别, 其二为自相似流量模拟。当前对网络自相似性分析的研究中, 关于网络自相似判别方法及自相似性程度研究较少, 有必要给出相关的自相似判别方法及其案例分析。当前常见的模拟自相似流量的方法有 ON/OFF 模型、M/G/∞ 排队模型、自回归分数滑动平均(Fractional AutoRegressive Integrated Moving Average, FARIMA)模型、小波模型及分形布朗运动(Fractal Brownian Motion, FBM)模型等, 其中, FBM 模型具有坚实的数学基础, 算法复杂度低, 模型参数少, 常被用来进行

长相关自相似流量模拟^[8]。随机中点置位(Random Midpoint Displacement, RMD)算法是产生 FBM 模型的理论基础, 但由于其在应用过程中会产生负值流量, 与实际情形相悖, 因此, 有必要对 FBM 模型进行改进, 以增强其对流量模拟的合理性。本文正是基于以上两个出发点, 首先给出一种自相似流量判别方法, 然后对 FBM 模型进行相关改进。

文献[9]提出的分形概念囊括了自相似性。分形数学是基于时间序列的矩分析, 通过分析样本矩的数学式, 可以得到序列分形特性。因此本文在已有分形矩分析基础上, 提出一种多阶矩的自相似判别方法。

分数布朗运动是一种统计自相似过程的数学模型, 主要用于生成布朗运动过程^[10]。1994 年 Norros^[11]首次通过 FBM 过程构建网络自相似业务流, 并给出该业务流的基本定义, 证明了其符合自相似性和长相关性。FBM 模型能够描述网络业务流的自相似特性, 只需要平均速率 v 、方差 α 和 Hurst 参数等三个参数就可以完整刻画整个模型, 其中 Hurst 参数是刻画网络的自相似程度。由于经典 RMD 算法会产生负值流量, 故本文对经典 RMD 算法进行相关改进以满足流量模拟的合理性。

收稿日期: 2012-10-12; 修回日期: 2012-11-30。

作者简介: 张雪媛(1978-), 女, 江苏泗洪人, 工程师, 硕士, 主要研究方向: 计算机网络、数据挖掘; 王永刚(1976-), 男, 江苏泗洪人, 讲师, 硕士, 主要研究方向: 计算机网络、指挥自动化; 张琮(1989-), 女, 河南驻马店人, 硕士研究生, 主要研究方向: 网络可靠性。

1 FBM 模型

1.1 自相似性

自相似是指局部结构与总体结构相比具有某种程度的一致性。自相似过程是一种统计意义上具有尺度不变性的随机过程。自相似性数学定义较多,其中标准的自相似过程定义^[9]如下:

定义 1 $Y = \{Y(t), t \in T\}$ 是一个连续时间过程的自相似过程,若它满足如下条件:

$$Y(t) \triangleq c^{-H} Y(ct); \forall t \in T, c > 0 \quad (1)$$

其中 H 是取值于 $(0.5, 1)$ 的 Hurst 参数,符号“ \triangleq ”表示任意有限维分布意义上相等。

定义 2 $\forall s, t \in T, R_Y(s, t) = E(Y(s)Y(t))$ 存在,则称 $R_Y(s, t)$ 为 $\{Y(t), t \in T\}$ 的自相关函数。

特别地,当 $\{Y(t)\}$ 是平稳过程,记

$$R_Y(k) = R(Y(s), Y(s+k)) \quad (2)$$

则自相似性过程中的 LRD 满足:

$$\sum_i R_Y(i) = \infty \quad (3)$$

且 H 越大, $R_Y(i)$ 衰减越慢,自相似程度也越高。

1.2 FBM 过程

FBM 过程 $\{X(t), t \in T\}$ 是一种典型的能产生自相似序列的过程,且满足:

$$R_X(s, t) = (s^{2H} + t^{2H} - |s - t|^{2H})/2 \quad (4)$$

为了更好地分析自相似性质,引入 m 阶聚合序列 $X^{(m)}$, 令:

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X(i); k = 1, 2, \dots \quad (5)$$

m 阶聚合序列通过将原始序列划分为大小为 m 的块,然后对每个块进行平均得到。

由式(1)可得:

$$X \triangleq m^{1-H} X^{(m)} \quad (6)$$

1.3 FBM 模型

网络流量 FBM 模型是基于 FBM 过程产生的一种流量模型。对于平均速率为 v , 方差系数为 $a(v, a > 0)$ 的网络流量 A_t , 令 Z_t 为自相似参数为 H 的 FBM 过程,网络流量累计模型(FBM 模型)可表示如下:

$$A_t = vt + \sqrt{av} Z_t; t \in (-\infty, +\infty) \quad (7)$$

在一般的分析中, A_t 的增量过程 ΔA_t 常表示为:

$$\begin{aligned} \Delta A_i(i) &= v + \sqrt{av}(Z_t(i+1) - Z_t(i)); \\ i &= \dots, -1, 0, 1, \dots \end{aligned} \quad (8)$$

对式(8)进行分析可知, FBM 模型形式较简单:一个均值加上一个带系数的 FBM 增量过程。这种模式已成为所有基于自相似过程的网络模型通式。FBM 模型在数学上有坚实的理论基础,因而可以很方便地应用于流量实时仿真和特性分析。

2 自相似流判别方法

自相似通常并不通过式(1)(即有限维分布)来研究,而是通过绝对矩的特性。对于随机过程 $\{X(t), t \in T\}$, 考虑

$$\mu^{(m)}(q) = E(|X^{(m)}|^q) \quad (9)$$

由文献[12]可知,若 $\{X(t)\}$ 是自相似,那么 $\mu^{(m)}(q)$ 正比于 $m^{\beta(q)}$, 即 $\mu^{(m)}(q) \propto m^{\beta(q)}$ 。

对于固定的 q , 有:

$$\log \mu^{(m)}(q) = \beta(q) \log m + C(q) \quad (10)$$

由式(6),可推导出

$$\beta(q) = q(H-1) \quad (11)$$

即式(8)可化为:

$$\log \mu^{(m)}(q) = q(H-1) \log m + C(q) \quad (12)$$

由式(9)和式(12)即可得到流量的自相似判别方法,方法流程如下所示:

1) 计算采样聚集序列 X 的 q 阶中心矩。

$$\hat{\mu}^{(m)}(q) = \frac{1}{N/m} \sum_{k=1}^{N/m} |X^{(m)}(k) - \frac{1}{N} \sum_{i=1}^N X(i)|^q \quad (13)$$

2) 改变 m 值,重新计算 $\hat{\mu}^{(m)}(q)$, 得到 m 序列和序列 $\hat{\mu}^{(m)}(q)$ 。

3) 判断 $\log \hat{\mu}^{(m)}(q)$ 序列和 $\log m$ 序列的关系:如果线性,则序列 X 自相似;否则,不自相似。

3 自相似流产生方法

根据式(8)可知,只要产生 FBM 过程序列,即可生成自相似序列 $\Delta A_t(i)$ 。文献[9]中采用标准正态过程特征值的方法模拟 FBM 过程的轨迹,但模拟产生的轨迹可能存在负值,不合实际。本文根据文献[13]中提出的经典 RMD 算法产生 FBM 过程,但经典 RMD 算法也存在该问题,即容易产生大的负值流量冲击波,不符合实际情况。

为了避免产生式(8)中差值为负的可能,本文对 RMD 产生的随机值进行从小到大的排序,具体如改进的 RMD 算法第4步所示。

改进的 RMD 算法生成 $[0, T]$ 内的 FBM 过程随机数步骤如下:

1) 定义 $Z_T(0) = 0$, $Z_T(T)$ 为服从均值为 0, 方差为 T^{2H} 正态分布的随机数。

2) 记 $Z_T(T/2) = (Z_T(0) + Z_T(T))/2 + I_1$, 其中 I_1 为均值为 0, 方差为 $[(1 - 2^{2H-2})/2^{2H}]T^{2H}$ 的服从正态分布的随机数。

3) 按照步骤2)中的方法,对于区间 $[a, b]$ 可得如下类推公式:

$$Z_T((a+b)/2) = (Z_T(a) + Z_T(b))/2 + I_k \quad (14)$$

其中: I_k 为服从均值为 0, 方差为 $I_{k-1}/2^{2H}$ 的正态分布随机数; k 为迭代次数。

4) 按照 1)~3) 的步骤产生了所需个数(个数依赖于迭代次数)的 FBM 过程的一个轨迹,从小到大排序得到 Z_1, \dots, Z_n , 该步避免了以前模拟网络流量出现增量为负的情形。

5) 选定要模拟的数据对象的平均速率 v 、方差 a 和 Hurst 参数,由式(8)即可得 FBM 模型的自相似随机序列。

4 案例验证

本文以 Bellcore 采集的局域网真实网络流量序列 BC-pAug89^[14] 和 LBL 实验室采集的广域网 LBL_PKT_5TCP^[15] 数据为例进行案例验证。首先通过第2章中提到的方法来分析这两列数据自相似性,然后以 BC-pAug89 数据为例,采用第3章方法产生仿真的自相似流量并与其进行对比分析。本文案例验证部分通过 Matlab 处理样本数据,并通过 Matlab-M 文件编写 RMD 程序,生成 FBM 过程的轨迹。

4.1 自相似流判别

根据式(13),首先确定 q , 计算 $\log m$ 与 $\hat{\mu}^{(m)}(q)$, 可分别得到 BC-pAug89 样本和 LBL_PKT_5TCP 样本 q 阶矩与聚类系数关系图,分别如图1~2所示,此处采集的是源数据的数据包到达间隔时间。

根据式(12)可知,若 $\log m$ 与 $\log \mu^{(m)}(q)$ 成线性关系,则该序列是自相似的。观察图1和图2可知均渐近成直线,表明 BC-pAug89 数据和 LBL_PKT_5TCP 数据都呈现出自相似性。根据文献[16]的方法提出的方法,即式(12)中, $\log m$ 与 $\log \mu^{(m)}(q)$ 构成的直线的斜率为 $q(H-1)$,从而可以计算出 BC-pAug89 数据的 Hurst 参数为 0.7469,该处得到的 Hurst 参数是平均意义,不同阶数得到的 Hurst 值不同,LBL_PKT_5TCP 数据的 Hurst 参数为 0.7865,后者的自相似程度更高。

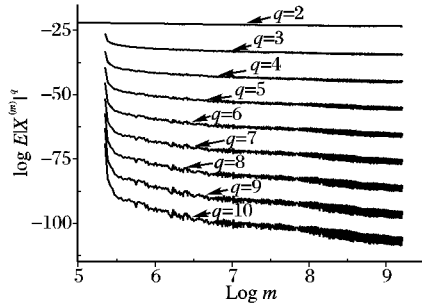


图1 BC-pAug89 样本 q 阶矩与聚类系数关系

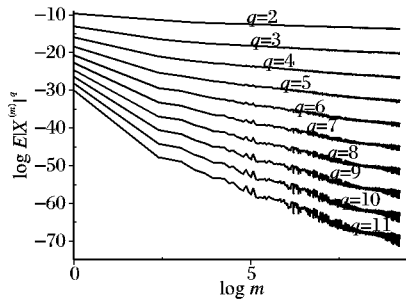


图2 LBL_PKT_5TCP 样本 q 阶矩与聚类系数关系

通过图1~2可以看到,不同阶数 q 对应不同的 Hurst 参数,这种现象可以通过自相似的尺度特性解释,其不同尺度下,自相似体现出局部的分形特性,也即 Hurst 参数随着尺度不同而呈现局部差异性。通过图1~2可知,基于多阶矩的自相似流量判别方法合理有效。

4.2 自相似流量模拟

本文选取 BC-pAug89 数据作为目标数据,通过 Matlab 统计得出平均速率 $\hat{\nu}$ 和方差 $\hat{\sigma}^2$,通过改进后的 RMD 算法实现 FBM 过程的一个轨迹,如图3所示。

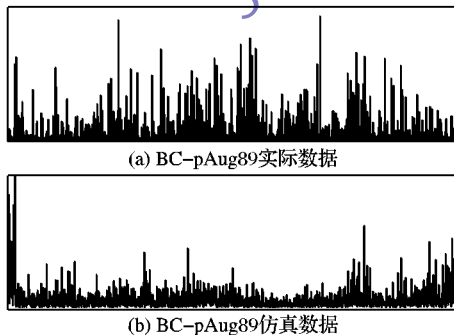


图3 实际仿真对比

从图3可以看到,模拟产生的数据具有良好的自相似性,同时体现了流量的总体稳定和局部突发性,对比实际数据具有较好的模拟效果。仿真模拟数据的 q 阶矩与聚类系数关系如图4。

从图4可以看到,模拟产生的数据具有良好的自相似性。按照图1~2计算 Hurst 参数的方法,可知仿真模拟数据的 Hurst 参数为 0.7335,较之原始数据的 0.7469 具有良好的替

代性,表明仿真数据的有效性。

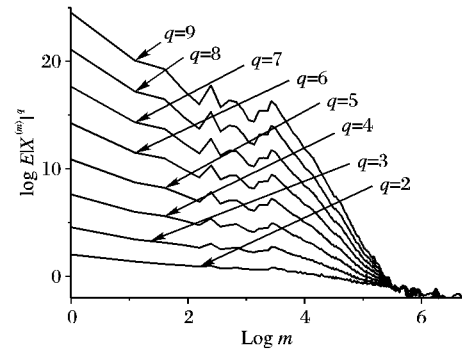


图4 仿真模拟数据 q 阶矩与聚类系数关系

同时从图3中也可以观察到,仿真得到的流量较之真实流量持续的突发性不强,这是由于自相似流量的尺度特性引起的,与4.1节 BC-pAug89 数据自相似判别分析对应,该数据在局部具有多分形性,即自相似程度在不同尺度下发生了改变,通过图3可知,改进后的 RMD 算法对流量的突发性进行了有效的仿真。

4.3 结果对比分析

通过图3的对比结果和图4,可以得到如下结论:

1)与经典 RMD 算法对比。文献[13]中的经典 RMD 算法能产生大量的负冲击波(见图5),直接可能导致模拟数据为负值,不合实际情况,而用本文改进的 RMD 算法则不会出现这种情况(图3(b)中没有大量的回落流)。因此,对比于经典 RMD 方法,本文方法得到的结果更符合实际。

2)与实际流量对比。由于 RMD 算法生成的流量保证了源数据流量的基本统计特征(均值和方差),同时通过仿真得到的 Hurst 参数也与真实源数据相近,验证了该仿真算法的有效性。

综合1)、2)结论可知,基于改进的 RMD 算法能有效实现长相关自相似流量模拟。

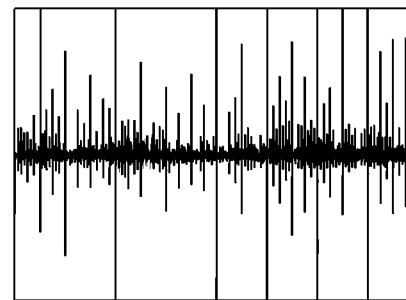


图5 经典 RMD 算法模拟产生的 BC-pAug89 流量

5 结语

流量模型是网络性能分析和通信网络规划设计的基础,精确的流量模型对设计高性能网络协议、业务量预测与网络规划、高性能网络设备与服务器、精确的网络性能分析与预测、拥塞管理都有重要意义。

本文首先通过 FBM 序列的 q 阶矩检验方法来评估序列的自相似性,并能根据对比 Hurst 参数评估不同序列的自相似性程度。其次,提出依据改进后的 RMD 算法生成 FBM 过程来仿真实际的自相似性序列。验证实验案例结果表明本文提出的判别和生成方法均具有良好实用性。经研究表明由于 FBM 模型是精确的自相似模型,且具有良好的数学特性,所以基于 FBM 过程的网络流量自相似模型判别和生成方法具有深远的实际应用意义。

(下转第963页)

然表现出了良好的跟踪性能,使得被控系统表现出良好的动态性能,为该算法今后在实际工业中的广泛应用奠定了基础。

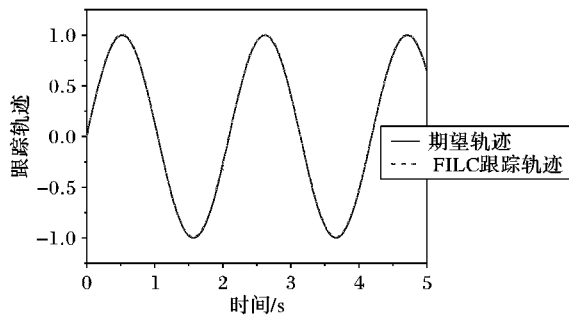


图10 关节1第5次跟踪轨迹

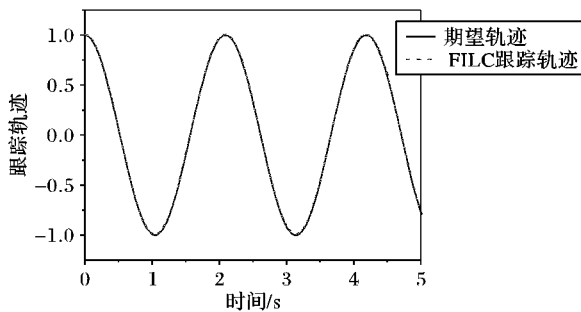


图11 关节2第5次跟踪轨迹

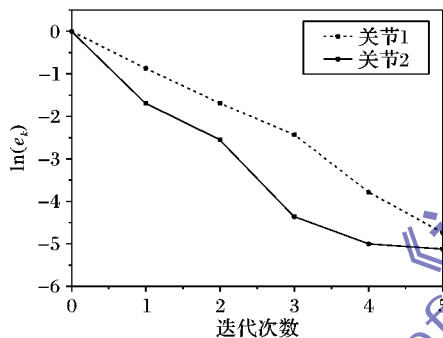


图12 双关节5次跟踪误差收敛曲线

4 结语

针对高精度控制轨迹的跟踪问题,本文提出一种基于遗

传算法的模糊迭代学习控制算法,学习算法由模糊 TSK 系统产生,TSK 模型参数由遗传算法整定,以达到精确跟踪和快速收敛的目的,并通过仿真验证了该方法的有效性,对同类控制对象的控制系统设计有一定的参考价值。

参考文献:

- [1] POK Y M, LIEW K H, XU J X. Fuzzy PD iterative learning control algorithm for improving tracking accuracy[C]// Proceedings of the 1998 IEEE International Conference on Systems, Man, and Cybernetics. Piscataway, NJ: IEEE Press, 1998: 1603 - 1608.
- [2] PRECUP R E, PREITL S, PETRIU E M. Iterative learning based fuzzy control system[C]// ROSE 2008: IEEE International Workshop on Robotic and Sensors Environments Proceedings. Piscataway, NJ: IEEE Press, 2008: 25 - 28.
- [3] 张丽萍, 杨富文. 基于 T-S 模型的非线性系统的迭代学习控制[J]. 系统仿真学报, 2009, 21(1): 199 - 203.
- [4] EBADAT A, KARIMAGHAE P, JESMANI M. Optimization-based fuzzy iterative learning control[C]// ICEE 2011: 2011 19th Iranian Conference on Electrical Engineering. Tehran, Iran: IEEE Computer Society, 2011: 1 - 6.
- [5] 陶诚, 黄圣国. 基于 TSK 模型的车辆跟驰模型[J]. 华南理工大学学报: 自然科学版, 2009, 37(3): 144 - 148.
- [6] 蒋思中, 朱芳来, 王心开, 等. 模糊增益 PD 型迭代学习算法及其应用[J]. 电光与控制, 2009, 16(8): 72 - 74.
- [7] 张丽萍, 杨富文. 基于 T-S 模型的迭代学习控制算法及其在机器人点位控制中的仿真研究[J]. 系统仿真学报, 2005, 17(1): 166 - 169.
- [8] 朱胜, 孙明轩. 具有位置死区输入非线性系统的迭代学习控制[J]. 控制与决策, 2009, 24(1): 96 - 100.
- [9] 晏静文, 侯忠生. 学习增强型 PID 控制系统的收敛性分析[J]. 控制理论与应用, 2010, 27(6): 761 - 768.
- [10] 赵丽莉, 孙明轩, 金奎. 一类非线性不确定系统的迭代学习控制[J]. 浙江工业大学学报, 2011, 39(2): 163 - 167.
- [11] 郝晓弘, 胡振邦, 王理平, 等. 一类模糊自适应 PID 迭代学习控制系统的研究[J]. 控制理论及其应, 2010, 32(5): 5 - 7.
- [12] 戴学丰, 严浙平, 孙立宁, 等. 双连杆柔性臂机器人的模糊补偿滑模控制[J]. 电机与控制学报, 2005, 9(1): 1 - 4.

(上接第 949 页)

参考文献:

- [1] LELAND W E, TAQQU M S, WILLINGER W, et al. On the self-similar nature of Ethernet traffic[J]. ACM SIGCOMM Computer Communication Review, 1993, 23(4): 202 - 213.
- [2] 高波, 张钦宇, 梁永生, 等. 基于 EMD 及 ARMA 的自相似网络流量预测[J]. 通信学报, 2011, 32(4): 47 - 56.
- [3] 徐兴, 岳学军, 林涛. 基于超爱尔兰分布的自相似网络排队模型研究[J]. 计算机工程与应用, 2012, 48(18): 71 - 75.
- [4] GRASSMANN W K, LUO J X. Simulating Markov-reward processes with rare events[J]. ACM Transactions on Modeling and Computer Simulation, 2005, 15(2): 138 - 154.
- [5] NOGUEIRA A, SALVADOR P, VALADAS R, et al. Markovian modeling of Internet traffic[C]// Network Performance Engineering, LNCS 5223. Berlin: Springer, 2011: 98 - 124.
- [6] BONALD T. The Erlang model with non-Poisson call arrivals[J]. ACM SIGMETRICS Performance Evaluation Review, 2006, 34(1): 276 - 286.
- [7] 高茜, 冯琦, 李广侠. 基于组合模型的自相似业务流量预测[J]. 计算机科学, 2012, 39(4): 123 - 126.
- [8] 张宾, 杨家海, 吴建平. Internet 流量模型分析与评述[J]. 软件学报, 2011, 22(1): 115 - 131.
- [9] MANDELBROT B B. The fractal geometry of nature[M]. [S. l.]: W. H. Freeman and Company, 1982.
- [10] 徐志强, 陈剑, 陆俊. 自相似混合业务流的网络性能研究[J]. 通信技术, 2012, 45(6): 117 - 120.
- [11] NORROS I. A storage model with self-similar input[J]. Queueing Systems, 1994, 16(3/4): 387 - 396.
- [12] TAQQU M S, TEVEROVSKY V, WILLINGER W. Is network traffic self-similar or multifractal[J]. Fractals, 1997, 5(1): 63 - 73.
- [13] LAU W C, ASHOK E, WANG J L, et al. Self-similar traffic generation: the random midpoint displacement algorithm and its properties[C]// Proceedings of ICC '95: IEEE International Conference on Communications. Piscataway, NJ: IEEE Press, 1995: 466 - 472.
- [14] Bellcore Morristown Research and Engineering Facility. BC[EB/OL]. [2011 - 08 - 29]. <http://ita.ee.lbl.gov/html/contrib/BC.html>.
- [15] Lawrence berkeley laboratory. LBL-PKT[EB/OL]. [2011 - 01 - 28]. <http://ita.ee.lbl.gov/html/contrib/LBL-PKT.html>.
- [16] KARAGIANNIS T, FALOUTOS M. SELFIS: a tool for self-similarity and long-range dependence analysis[EB/OL]. [2011 - 07 - 23]. <http://www.cs.ucr.edu/~tkarag/Selfis/Selfis.html>.