

# 网络流量的联合多重分形模型及特性分析

魏进武, 邬江兴, 陈庶樵

(国家数字交换系统工程技术研究中心, 河南郑州 450002)

**摘要:** 网络尺度行为的发现提供了用数学模型方法研究网络流量特性的可能性. 本文基于连乘瀑布过程与 K 分布过程提出了联合多重分形 (JMF) 网络流量模型, 该模型以尺度函数与矩因子的联合作为主要特征函数来研究网络流量的特性. 理论分析及由实测网络流量数据的仿真结果表明, JMF 模型可以较客观地同时描述网络流量短期分形行为与长期自相似行为, 且实现复杂度小. 其中尺度函数能够刻画时间尺度对流量特性的影响, 矩因子描述了同一时间尺度上流量突发性的变化, 二者的联合较好地描述了网络流量的短期行为, 而模型的统计特性则刻画了流量的长期行为特征.

**关键词:** 因特网; 网络流量; 多重分形; K 分布

**中图分类号:** TN271

**文献标识码:** A

**文章编号:** 0372-2112 (2004) 09-1459-05

## Joint Multifractal Model and Characteristics Analysis of Network Traffic

WEI Jin-wu, WU Jiang-xing, CHEN Shu-qiao

(National Digital Switching System Engineering & Technology R&D Center, Zhengzhou, Henan 450002, China)

**Abstract:** The discovery of scaling behavior in network has provided hope that mathematical models can be used to describe the nature of the traffic. A joint multifractal (JMF) model for network traffic is proposed based on the combination of multiplicative cascade process and independent K distributed process. The model focuses on both the scaling function and moment factor as the key for unraveling the multifractal characteristics of the network traffic. JMF model is objective and simple enough to describe the short-term fractal behavior as well as the long-term self-similar behavior of network traffic simultaneously, which is proved by theoretical analysis and simulated results from several real traffic data sets. The influence on the traffic characteristics due to timescale is expressed by scaling function, and traffic burstiness variation in a specified timescale is illuminated by moment factor. The combination of scaling function and moment factor describes the short-term behavior of network traffic, and the statistical properties of JMF model describe the traffic long-term behavior characteristics.

**Key words:** Internet; network traffic; multifraction; K distribution

### 1 引言

网络流量模型及流量特性分析对于网络及网络节点优化设计与仿真, 提供服务质量 (QoS) 保证以及网络管理与控制等方面都具有重要的意义. 已有大量的网络流量模型<sup>[1-5]</sup>试图刻画网络流量的变化特性, 但仍难以客观地描述流量复杂的尺度行为. 本文的主要目的是建立一个能够客观刻画流量特性且易实现的网络流量模型, 并通过理论证明与数据仿真的方法来分析其特性.

大量的研究结果以及测试数据表明实际网络流量在不同的时间尺度上显示出剧烈的变化与突发特性<sup>[6-9]</sup>, 该发现促使当前大量的研究都集中于如何准确地捕获网络流量的长

关 (LRD) 特性及如何准确高效地估计 Hurst 指数<sup>[9]</sup>, 例如混沌图<sup>[1]</sup>, 长相关的 ON/OFF 模型<sup>[2]</sup>, 基于排队过程的 M/G 类模型<sup>[3]</sup>, 分形布朗运动 (FBM) 模型<sup>[4]</sup> 以及分形自回归平滑 (FARIMA) 模型<sup>[5]</sup> 等. 在这些模型中, FARIMA 模型能够较好地描述 LRD 特性, 但是其运算量大, 一般产生长度为 N 的合成流, 需要的运算复杂度为  $O(N^2)$ , 虽然 C Bor-Sen 等人试图用遗传算法<sup>[5]</sup> 解决该问题, 但并没有明显改善. FBM 模型由于其高斯分布特性而被广泛应用, 但用高斯分布研究非均匀网络流量的特性显得不足. 总之, 这些模型能够较好地刻画网络流量的长期行为特征, 但对其短期行为的细微特征描述方面显得不够充分. 而且流量数据的分析结果表明, 自相似框架下的流量模型不能充分研究流量复杂的尺度行为<sup>[10]</sup>. M Sheng 等

人在小波域中构建了网络流量的马尔可夫小波模型<sup>[7,8]</sup>,其研究结果表明,具有长相关特性的网络流量对应的小波系数是短相关的,在大时间尺度上,网络流量是自相似过程,而在小时间尺度上,网络流量具有多重分形特性,且呈非高斯分布.对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,能表达其细微特征.对于流量控制,了解短时间内流量的突然变化要比了解其长期相关性更为重要,并且对于非均匀的流量过程,单维数是无法描述其全部特征的.多重分形能全面考察流量的突发性及其变化,以此方法研究网络流量更为适合.

本文基于连乘瀑布模型<sup>[10]</sup>与独立的 K 分布<sup>[12]</sup>过程提出了网络流量的联合多重分形(JMF)模型,该模型综合考虑了流量过程的尺度函数与矩因子,因为尺度函数刻画了时间尺度对流量特性的影响,而矩因子则描述了流量突发性相对时间的变化.尺度函数与矩因子的联合能够同时刻画网络流量的短期变化特性,而其长期行为可由模型在整个时域上的统计特性来描述.文中提出了尺度函数及矩因子的对数图联合估计方法.为了验证模型的有效性,及研究真实 WAN 流量的特性,分别对 1995 年采自劳伦斯伯克利国家实验室(LBNL)的两组数据<sup>[11]</sup>以及 2003 年由美国国家网络应用研究实验室(NLANR)采集的一组数据<sup>[13]</sup>进行了分析.本文第二节给出了 JMF 网络流量模型及其参数提取方法,第三节对其统计特性进行了理论分析,第四节对 3 组实测数据进行仿真分析,最后给出了本文的主要结论.

## 2 JMF 网络流量模型

根据文[6]的定义,若一个随机过程  $X(t), t > 0$ , 且增量满足  $E[X(t)|^q] = c(q) t^{(q)+1} = c(q) t^{o(q)}$  (1) 称其为分形过程.其中  $q \in Q, [0, 1] \subset Q, o(q)$  为尺度函数,  $c(q)$  为矩因子,  $X(t)$  为时间增量为  $t$  时的网络流量(包或者字节).若  $o(q)$  是  $q$  的线性函数,则该过程为单分形过程,否则若  $o(q)$  是  $q$  的凸函数时,该过程为多重分形过程.

### 2.1 模型建立

假设连乘瀑布过程的乘数因子  $R$  服从  $p_R(R)$  分布,则  $o(q)$  为<sup>[11]</sup>

$$o_2 E[R^q] \quad (2)$$

为了能够得到尺度函数  $o(q)$  及矩因子  $c(q)$  的联合信息,我们假设由乘数因子服从  $p_R(R)$  分布的连乘瀑布过程产生长度为  $2^N$  的数据序列  $(t_N)$ ,将其与相同长度的随机序列  $\{S_k \geq 0, k = 1, 2, \dots, 2^N\}$  相乘,序列  $\{S_k\}$  由独立同分布(I.I.D)的随机变量  $S$  构成并且与  $(t_N)$  相互独立,则

$$E[X(t_N)|^q] = E[S^q] E[(t_N)^q] = E[S^q] t_N^{o(q)} \quad (3)$$

在实际应用中,我们对此做一些修正.  $(t_N)$  是  $N$  个乘数因子  $R$  的乘积,为了避免由于  $(t_N)$  过小而造成信息损失,我们将其与长度因子  $2^N$  相乘,使  $2^N (t_N)$  具有单位期望值.另外,对瀑布过程进行重新标度,使其在第  $N$  阶段具有单位时间间隔,而不是原始的  $t_N = 2^{-N}$ .将修正内容代入式(3),可得

$$E[X(t)|^q] = E[S^q] 2^{N(q+log_2 E[R^q])} t^{-log_2 E[R^q]} \quad (4)$$

其中  $X(t)$  表示单位时间间隔内网络的流量.对比式(4)与多重分形的定义式(1)发现,建模的任务就是选择合适的随机变量  $R$  与  $S$ ,使得

$$\begin{cases} -\log_2(E[R^q]) = o(q) \\ E[S^q] 2^{N(q+log_2 E[R^q])} = c(q) \end{cases} \quad (5)$$

如果实测的网络流量数据具有分形特性,则可由尺度函数  $o(q)$  及矩因子  $c(q)$  来模拟网络流量.

### 2.2 模型参数提取

JMF 模型需要联合估计尺度函数及矩因子,本文提出了一个关于  $o(q)$  及  $c(q)$  联合估计的算法—对数图法.根据式(1)对多重分形过程的定义,有

$$\log_2 E[X(m-t)|^q] = o(q) \log_2(m-t) + \log_2 c(q) \quad (6)$$

其中  $m = 1, 2, \dots, Z(m-t)$  表示实测网络流量在某时间尺度上的第  $m$  层聚合.基于这一特性,假设其增量为  $Z(t), Z(2t), \dots, Z(m-t)$ ,定义其相应的聚合序列  $\{Z(m-t)\}$  为

$$Z_k(m-t) = \prod_{i=km-(m-1)}^{km} Z_i(t) \quad (7)$$

其中  $k, m = 1, 2, \dots$  如果  $\{Z(m-t)\}$  具有尺度特性,则在对数图上,其矩  $\log_2 E[X(m-t)|^q] = \log_2 m$  与将呈线性关系,如图 1 所示,则其斜率为  $o(q)$ ,截距为  $\log_2 c(q)$ .文中,由实测数据中估计的尺度函数及矩因子分别用  $\hat{o}_0(q)$  及  $\hat{c}(q)$  来表示.

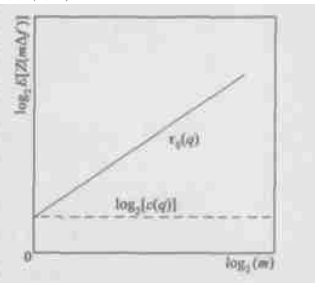


图 1 估计尺度函数及矩因子的原理图

文献[10]表明,  $[0, 1]$  上的对称 Beta( $\alpha, \beta$ ) 分布在  $\alpha, \beta > 0$  时可以较好的拟合  $\hat{o}_0(q)$ , 参数  $\alpha, \beta$  满足

$$\log_2 \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)} = \hat{o}_0(q) \quad (8)$$

其中  $\Gamma(\cdot)$  为 gamma 函数.

由于 K 分布适合描述奇异现象及事件的突发行为,本文选择独立 K 分布<sup>[12]</sup>的随机变量  $S$ . 设  $S$  具有参数  $(\nu, \mu)$ , 则  $S$

的  $q$  阶矩为  $E[S^q] = \left(\frac{2\nu}{\mu}\right)^{q/2} \frac{\Gamma(\nu + \frac{q}{2}) \Gamma(\frac{q}{2} + 1)}{\Gamma(\nu)}$ . 比较式(1)与式(5)知,  $(\nu, \mu)$  的选择应满足

$$\left(\frac{2\nu}{\mu}\right)^{q/2} \frac{\Gamma(\nu + \frac{q}{2}) \Gamma(\frac{q}{2} + 1)}{\Gamma(\nu)} = \hat{c}(q) 2^{N(\log_2 \frac{\Gamma(\nu + \frac{q}{2}) \Gamma(\frac{q}{2} + 1)}{\Gamma(\nu)} - q)} \quad (9)$$

由此, JMF 流量模型仅需  $(\nu, \mu)$  3 个参数, 并且参数要满足

$$\begin{cases} \hat{o}_0(q) = \log_2 \frac{\Gamma(\nu + \frac{q}{2}) \Gamma(\frac{q}{2} + 1)}{\Gamma(\nu)} \\ \hat{c}(q) = \left(\frac{2\nu}{\mu}\right)^{q/2} \frac{\Gamma(\nu + \frac{q}{2}) \Gamma(\frac{q}{2} + 1)}{\Gamma(\nu)} 2^{N(q - \log_2 \frac{\Gamma(\nu + \frac{q}{2}) \Gamma(\frac{q}{2} + 1)}{\Gamma(\nu)})} \end{cases} \quad (10)$$

### 3 JMF 流量模型的统计特性

本节分析 JMF 流量模型的统计特性. 对于由 JMF 模型产

生的长度为  $2^N$  的网络流量,具有:

(1)由多重分形的定义及建模过程可知,尺度函数  $\phi(q)$  及矩因子  $c(q)$  的联合共同表征了 JMF 模型的多重分形特性;

(2)JMF 模型得到的合成过程  $X(t)$  的矩特性分别为

$$E[X(t)] = E[S] = \sqrt{\frac{v}{2\mu}} \frac{(v + \frac{1}{2})}{(v)} \quad (11)$$

$$\begin{aligned} \text{var}[X(t)] &= E[S^2] 2^{2N} E[R^2]^N - E[X(t)]^2 \\ &= \frac{2v^2}{\mu} \left( \frac{v+1}{v+1/2} \right)^N - \frac{v^2 (v + \frac{1}{2})}{2\mu^2 (v)} \end{aligned} \quad (12)$$

(3)如果  $N \gg 1$ ,即样本足够多的情况下,合成的网络流量  $X(t)$  服从 K 分布,通过连乘瀑布过程的统计特性<sup>[10]</sup>及大数定理可以证明.样本足够多描述的是流量的长期行为为特征,即 K 分布可以描述网络流量长期行为的统计分布特性;

(4)  $X(t)$  具有长相关特性.因为其协方差函数可表示为

$$\begin{aligned} \text{cov}[X(t)_n, X(t)_{n+k}] &= E[S]^2 \{ 2^{2N} E[(t_N)_n, (t_N)_{n+k}] - 1 \} \end{aligned} \quad (13)$$

根据文献[10]推导的结果,连乘瀑布过程  $(t_N)$  的相关函数为

$$E[(t_N)_n, (t_N)_{n+k}] = 2^{-2p} E[R^2]^{N-p-1} (1/2 - E[R^2]),$$

其中  $k = 2^p, p = 1, 2, \dots$ , 则

$$\text{cov}[X(t)_n, X(t)_{n+k}] = \frac{2v^2}{\mu} \left( \frac{v+1}{v+1/2} \right)^{N-1} k^{-\log_2 \left( \frac{v+1}{v+1/2} \right)} - \frac{2v^2}{\mu} \quad (14)$$

由此可见如果在  $N, k$  足够大时,即样本足够多的情况下,合成流量的协方差函数的变化情况完全由因子  $k^{-\log_2 \left( \frac{v+1}{v+1/2} \right)}$  决定,根据自相似过程的定义,则 Hurst 指数  $H = 1 - \frac{1}{2} \log_2 \left( \frac{v+1}{v+1/2} \right)$ ,显然对于所有  $v > 0$ ,均有  $H \in (0.5, 1)$ ,表明网络流量的长期行为具有长相关特性.

以上对 JMF 流量模型所产生流量统计特性的分析表明, JMF 流量模型既能通过尺度函数及矩因子的联合来刻画网络流量短期行为上的多重分形特性,又可以通过其统计特性较好地刻画流量长期行为的长相关特性,并且 K 分布可以描述流量长期行为的统计分布特性.

### 4 数据分析及仿真实验

本节中分别对 1995 年采集自 LBNL 的两组数据 DEC-PKF3 与 DEC-PKF4<sup>[11]</sup>以及 2003 年在 NLANR 采集的一组数据<sup>[13]</sup>TXS-1047278559 进行了分析,并由此提取相应的尺度函数及矩因子,以检验 JMF 流量模型对全面描述网络特性的有效性. DEC-PKF3 与 DEC-PKF4 数据是用 tcpdump 捕获的 1 小时内所有数据包的 IP 头, TXS-1047278559 数据由 NLANR 实验室在 GgaPOP 口采集的德州大学的网络流量数据,文中记为 TXS.

#### 4.1 真实网络流量数据分析

在我们分析的过程中,网络流量可以以不同时间尺度下的包及字节数为单位进行分析.图 2,图 3 与图 4 分别给出了 DEC-PKF-3, DEC-PKF-4 及 TXS 在 0.1 秒及 1 秒时间尺度下用

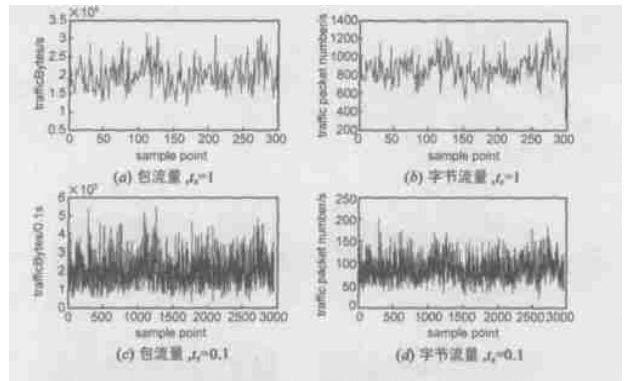


图 2 不同时间尺度下 DEC-PKF3 流量

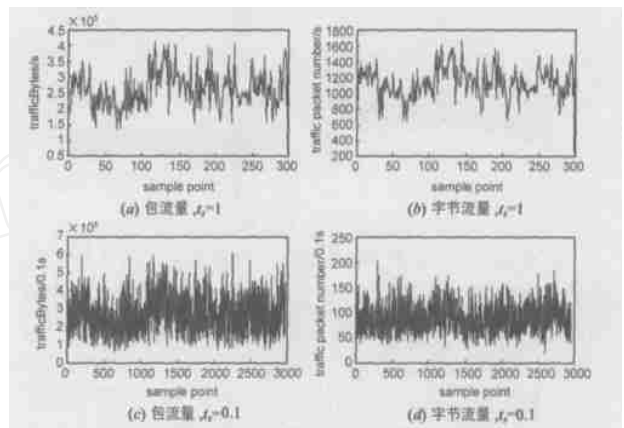


图 3 不同时间尺度下 DEC-PKF4 流量

包与用字节数为单位表示的网络流量,其中  $t_s$  代表时间尺度.图 2,图 3 以及图 4 所示的不同时间尺度下流量的时域曲线表明,不同的时间尺度下,网络流量的起伏特性大体上相同,但在某些采样点上,随着聚合程度的增加,其起伏特性(即突发性)有较明显的变化.例如图 4(c)及图 4(d)所示的 0.1 秒尺度下的流量特性相对于图 4(a)及图 4(b)所示的 1 秒尺度下的流量,其起伏程度更趋向于均匀.另外,在相同时间尺度下,用包表示的流量及用字节表示的流量的起伏特性基本相似,例如图 2(a)与图 2(b),即包流量与字节流量基本上呈线性变化.三组数据的字节流量及包流量的比较如图 5 所示.

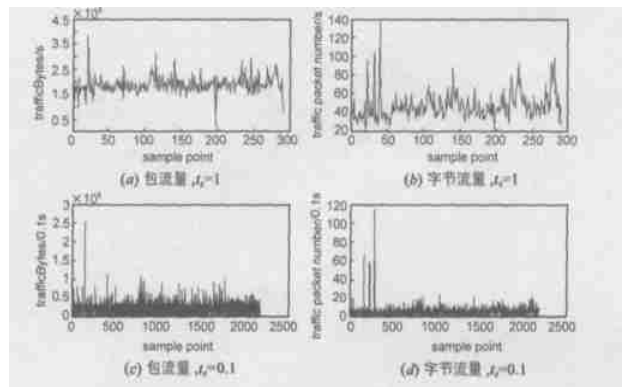


图 4 不同时间尺度下 TXS 流量

用“·”表示的包流量与用“x”表示的字节流量的汇聚性较为吻合.由此比较,流量的时域表示,仅能表达包到达时间对网络流量的突发性的影响,而包大小的随机分布对其影响在时域曲线上并不能非常明显的得到反映,即时域特性不足以较好反映流量的特性.

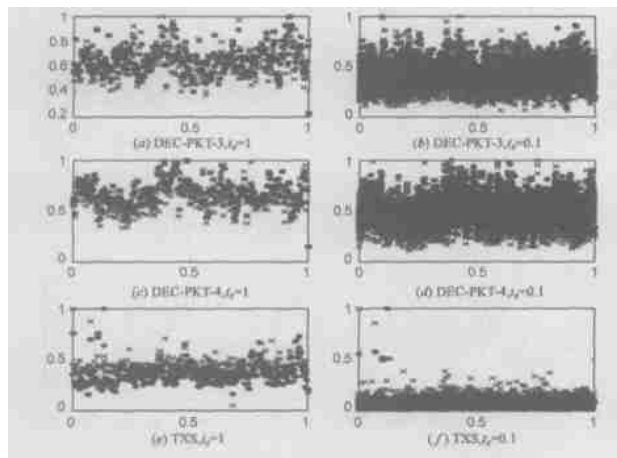


图5 包与字节表示的流量比较(·包, x 字节)

图6给出了用对数图法提取的三组流量数据在不同时间尺度下的尺度函数及矩因子的变化趋势.时间尺度对流量特性的影响可由图6(a)所示的尺度函数表示,随着聚合程度 $q$ 的增加,不同时间尺度下同一组数据的尺度函数差异越大,即其特性越明显.例如DEC-PKT-4在0.1秒时间尺度下的包流量与在1秒时间尺度下的包流量的尺度函数具有明显的变化,说明尺度函数能够很好地描述时间尺度对流量特性的影响,可作为描述网络流量短期行为特性的一个重要特征.此外结合多重分形的定义,DEC-PKT3仅表现出自相似特性,即其对应的尺度函数呈线性变化.而DEC-PKT4及TXS则表现出明显的多重分形特性.图6(b)给出的矩因子明显显示了包长对流量特性的影响,例如DEC-PKT-4在1秒尺度下包流量及字节流量对应的矩因子有明显差异,并且随着聚合程度的增加,矩因子的差异增大.由此,说明矩因子能够明确表示相同时间尺度上流量突发性的变化程度.通过比较分析表明,相对图2~4所示的流量时域特性而言,图6所示的尺度函数与矩因子的联合,可以在联合域中较全面地刻画网络流量短期行为的细微特征.

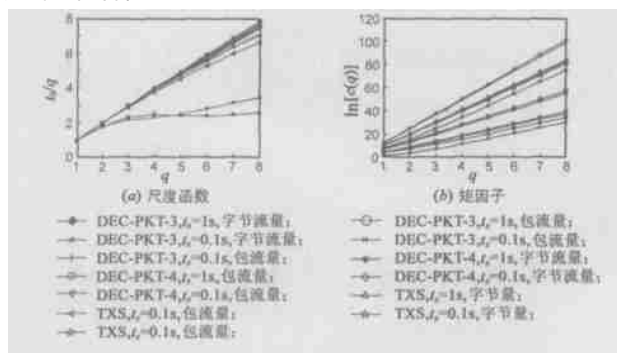


图6 不同尺度下三组实测数据的尺度函数及矩因子

### 4.2 JMF 流量模型的数据拟合

以TXS数据为例,我们通过模型中的尺度函数及矩因子提取JMF流量模型所需的参数,并对JMF模型进行拟合比较,所选用的时间尺度为1秒,对其1小时所采集的数据进行了拟合.通过以上对尺度函数的研究(图6(a)),可以计算出以字节为单位表示的流量对应的瀑布过程中的乘数因子所服从的Beta分布的参数 $\nu_1 = 19.245$ ,用包表示的流量对应的参数 $\nu_2 = 12.023$ .通过矩因子可计算出独立K分布对应于字节流量的参数集分别为 $\nu_1 = 5.6611, \mu_1 = 0.653$ ,与包表示的流量对应K分布的参数集 $\nu_2 = 2.3411, \mu_2 = 0.063$ .真实流量数据估计得到的尺度函数及矩因子与JMF模型所得到结果如图7所示.通过比较,发现二者均能紧密吻合,考虑到尺度函数与矩因子的联合可较客观地描述流量特性,故说明JMF流量模型能够较好地拟合真实网络流量.

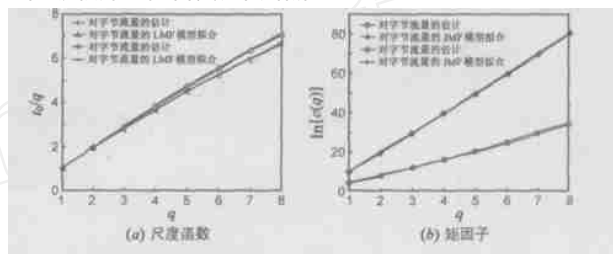


图7 对时间尺度为1秒TXS包及字节流量的拟合

### 5 结论

本文研究了网络流量建模及其特性问题,在连乘瀑布过程及独立K分布过程的基础上,提出了网络流量的JMF模型.通过理论分析及计算机仿真的方法研究了网络流量的多重分形特性及模型的统计特性.考虑到长相关仅仅是描述流量长期行为的一个特征,不足以充分刻画流量短期行为所具有的细微特征,JMF流量模型通过网络流量的尺度函数及矩因子的联合及其统计特性来分析流量短期及长期行为特性.理论分析结果表明,尺度函数及矩因子的联合能够全面研究网络流量的短期分形行为特征,而模型的统计特性则能够表达网络流量长期的长相关特性,K分布能够较好地拟合网络流量长期行为的统计分布特性.

三组实测数据的仿真分析表明,尺度函数能够明确反映时间尺度对流量特性的影响,随着聚合程度的增加,尺度函数随时间尺度的变化增大.此外,相同时间尺度上的包流量与字节流量的矩因子具有较明显的差异,并且随着聚合程度的增加而增大,表明矩因子能够描述相同时间尺度上流量突发性的变化程度.对1小时内数据网络实测数据进行了尺度函数及矩因子的估计,其结果与模型拟合的结果紧密吻合,表明JMF模型可以描述网络流量的长期与短期行为特性,刻画数据网络流量特性.

### 参考文献:

[1] A Erramilli, P R Singh, P Pruthi. An approach of deterministic chaotic maps to model packet traffic[J]. Queueing Systems, 1995, 20(3): 171 -

- 206.
- [ 2 ] W Willinger ,M S Taqqu ,M S Sherman ,D V Wilson. Self-similarity through high-variability :Statistical analysis of ethernet LAN traffic at the source level [J ] . IEEE/ ACM Transactions on Networking ,1997 ,5 (1) :71 - 86.
- [ 3 ] B Tsybakov ,N D Georganas. On self-similar traffic in ATM queue :Definitions ,overflow probability bound ,and cell delay distribution [J ] . IEEE/ ACM Transactions on Networking ,1997 ,5 (2) :397 - 409.
- [ 4 ] J M Bardet. Statistical study of the wavelet analysis of fractal Brownian motion [J ] . IEEE Transactions on Information Theory ,2002 ,48 (4) :991 - 999.
- [ 5 ] C Bor-Sen ,L Bore-Kuen ,P Sen-Cheuh. Maximum likelihood parameter estimation of F-ARIMA processes using the genetic algorithm in the frequency domain [J ] . IEEE Transactions on Signal Processing ,2002 ,50 (9) :2208 - 2220.
- [ 6 ] D D Trang ,M S ador ,M Istv an. Some result on multiscale queuing analysis [A ] . Conf on ICT2003 [ C ] . Papeete ,French Polynesia : ICT ,2003.
- [ 7 ] M Sheng ,J Chuanyi. Modelling heterogeneous network traffic in wavelet domain [J ] . IEEE/ ACM Transactions on Networking ,2001 ,9 (5) :634 - 649.
- [ 8 ] T Xusheng ,M Sheng ,J Chuanyi. A unified framework for understanding network traffic using independent wavelet models [A ] . IEEE INFOCOM [ C ] . New York ,USA :IEEE ,2002.
- [ 9 ] W E Leland ,M S Taqqu ,W Willinger ,D V Wilson. On the self-similar nature of ethernet traffic (Extended version) [J ] . IEEE/ ACM Transactions on Networking ,1994 ,2 (1) :1 - 15.
- [ 10 ] A C Gilbert ,W Willinger ,A Feldmann. Scaling analysis of conservative cascades ,with applications to network traffic [J ] . IEEE Transactions on Information Theory ,1999 ,45 (3) :971 - 991.
- [ 11 ] Peter ,Dan Zig ,et al. The Internet traffic archive [DB/ OL ] . <http://ita.ee.lbl.gov>.
- [ 12 ] F Berlzi ,E Dalle Mese ,G Pinelli. One-dimensional fractal model of the sea surface [J ] . IEE Proc-Radar Sonar Navig ,1999 ,146 (1) :55 - 64.
- [ 13 ] Steven Paxson NLANR network traffic packet header traces [OB/ OL ] . <http://pma.nlanr.net/Traces/>.

### 作者简介:



**魏进武** 男,1978 年出生于甘肃庆阳,2002 年毕业于空军雷达学院信号与信息处理专业,获工学硕士学位,现为信息工程大学通信与信息工程专业博士研究生,目前的主要研究方向为 Internet 网络流量分析及下一代网络体系结构。



**邬江兴** 男,1953 年出生于安徽金寨,工程院院士,博士生导师,现为信息工程大学国家数字交换系统工程技术研究中心主任,主要的研究方向为信息网络与交换技术。



**陈庶樵** 男,1973 年出生于陕西西安,2002 年毕业于信息工程大学通信与信息系统专业,获工学博士学位,已发表学术论文 10 余篇,主要的研究领域为 Internet 服务质量保证及高端路由器体系结构。