

自相似网络流量差分分析新方法^{*})

李立 喻莉 朱光喜 白云

(华中科技大学电子与信息工程系/武汉光电国家实验室 武汉 430074)

摘要 多媒体网络流量具有重尾特性及自相似特性,已得到了广泛的认同。而 Alpha-stable 过程作为 Gaussian 过程的推广能很好地描述重尾特性及自相似特性。在简要介绍了 Alpha-stable 分布和 Alpha-stable 过程的基础上,本文研究分析了当前基于线形分形稳定噪声(LFSN)理论的网络业务流量模型,并提出了基于差分分析的新方法及基于差分分析的 LFSN 流量模型参数估计新方法。对 Bellcore 实验室采集数据的拟合验证结果表明,该估计方法能准确确定模型参数。

关键词 网络流量建模, Alpha-stable 分布,线形分形稳定噪声,差分分析,参数估计

New Analysis Method for Self-similar Network Traffic Based on the Differential Analysis

LI Li YU Li ZHU Guang-Xi BAI Yun

(Department of Electronics & Information Engineering, Huazhong University of Science and Technology, Wuhan 430074)

Abstract The fact that multi-media network traffic is with heavy-tail character and self-similar character has been known to researchers. At the same time, Alpha-stable process, the extension of Gaussian process, has the advantage of modeling heavy-tail character and self-similar character. Alpha-stable distribution and Alpha-stable process are introduced at the beginning of this paper. Then after analyzing of current network traffic models based on Alpha-stable process and Linear Fractional Stable Noise (LFSN), we advance a differential analysis method and a new parameter estimation method based on difference analysis. Results of analysis of actual traffic data from Bellcore show the method can get parameters precisely.

Keywords Network traffic model, Alpha-stable distribution, Linear fractional stable noise, Difference analysis, Parameter estimation method

1 引言

W. E. Leland, W. Willinger 等人^[1]分析了 Bellcore 实验室收集的大量以太网实际数据流量,并进行研究,结果发现以太网实际网络流量呈现出完全不同于传统泊松模型的特性:重尾特性(Noah 效应)和自相似性(Joseph 效应)。此后,研究人员又陆续对广域网数据流量^[2]、VBR 视频业务流^[3]等实际流量进行了研究分析,进一步证实了无论网络拓扑结构、服务类型等因素如何变化,网络流量的重尾特性和自相似性始终存在。

具有符合 Gaussian 分布平稳增量的自相似过程——分形布朗运动(Fractional Brownian Motion, FBM),是最早被用来对自相似网络流量建模的模型之一。但平稳增量服从 Gaussian 分布,使得 FBM 缺乏描述高突发性的多媒体网络流量的能力。针对这个问题,文[4]给出的结论说明了,许多有着严格交替周期的、独立且同分布的 ON/OFF 源,当它们表现 Noah 效应时,其叠加导致了网络流量的自相似性,并且其极限过程为 Levy 稳定运动。另一方面,文[5~7]讨论了基于 Alpha-stable 分布的自相似过程——线性分形稳定噪声(Linear Fractional Stable Noise, LFSN)的网络流量模型。与 Gaussian 分布相比,Alpha-stable 分布具有更多的自由度,以描述网络流量的突发性,但是也为理论研究带来很多困难。在分析当前一类基于 LFSN 的网络业务流量模型的基础上,本文利用差分分析方法提出了一般化 LFSN 自相似网络流量

模型参数估计新方法。

本文分为如下几个部分:第 2 节简要介绍 Alpha-stable 分布及 LFSM 的相关定义;第 3 节在分析现有基于 LFSM 的自相似网络流量模型的基础之上,提出一种新流量差分分析方法并在此基础上建立一般化 LFSM 自相似网络流量模型参数估计方法;第 4 节利用模型参数估计新方法对实际网络业务流量进行了分析;最后为全文总结。

2 Alpha-stable 分布和 LFSN 过程

2.1 Alpha-stable 分布的定义

由于 Alpha-stable 分布不具有闭形式的概率密度分布函数(Probability Density Function, PDF),但它能被特征函数(Characteristic Function, CF)唯一描述。标准参考系中定义的 Alpha-stable 分布 $X \sim S_{\alpha}(\sigma, \beta, \mu)$ 特征函数为^[9]

$$\Phi(\theta) = E \exp i\theta X =$$

$$\begin{cases} \exp\left\{-\sigma |\theta|^{\alpha} \left(1 - i\beta(\operatorname{sign}\theta) \tan \frac{\pi\alpha}{2}\right) + i\mu\theta\right\}, \alpha \neq 1 \\ \exp\left\{-\sigma |\theta| \left(1 + i\beta \frac{2}{\pi}(\operatorname{sign}\theta) \ln|\theta| + i\mu\theta\right)\right\}, \alpha = 1 \end{cases}$$

其中 α 为特征因子(Characteristic Exponent), σ 为尺度参数(Scale Parameter), β 为偏斜因子(Skewness Parameter), μ 为位移参数(Shift Parameter)。参数取值范围为: $0 < \alpha \leq 2, \sigma \geq 0, -1 \leq \beta \leq 1, \mu \in (-\infty, \infty)$ 。

2.2 线形分形稳定噪声(LFSN)

如果随机过程 $\{L_{S,H}(t), t > 0\}$ 满足下列条件,则称之为

^{*})国家自然科学基金资助项目(60502023)。李立 博士生,主要研究方向为计算机网络和多媒体通信;喻莉 博士,教授,博士生导师,主要研究方向为计算机网络和多媒体通信;朱光喜 教授,博士生导师,主要研究方向为多媒体和 B3G 移动通信计算机、图像处理;白云 博士生,主要研究方向为计算机网络和多媒体通信。

线性分形稳定运动:

$$L_{S,H}(t) = \int_{-\infty}^{\infty} (|t-x|^{H-1/\alpha} - |x|^{H-1/\alpha}) M_s ds$$

其中: $0 < \alpha < 2, 0 < H < 1, H \neq 1/\alpha, M_s$ 为 Alpha-stable 随机测度。

LFSM 的平稳增量过程为线性分形稳定噪声 (LFSN) 过程, 其离散近似表达式为^[9]

$$\hat{N}_{\alpha,H}(i) = (h_d * S_{\sigma,\beta,0}^{\alpha})(i) = \sum_{k=1}^{Km} h_d(k/m) S_{\sigma,\beta,0}^{\alpha}(i-k/m) \quad (1)$$

其中: $h_d(x) = \begin{cases} x^d - (x-1)^d, & 1 < x \\ x^d, & 1 < x \leq 1 \end{cases}, d = H - 1/\alpha, m$ 为积分离散化中的网格参数, k 为积分截断点。

3 基于 LFSN 的网络流量增量的模型

3.1 基于 LFSN 的网络流量增量模型

由(1)式可以得到基于 LFSN 过程的网络流量模型的如下基本形式:

$$\hat{M}(i) = c_1 \hat{N}_{\alpha,\beta,H}(i) + c_2 = c_1 (h_d * S_{\sigma,\beta,0}^{\alpha})(i) + c_2$$

其中: $\hat{M}(i)$ 为单位时间内的网络流量; $\hat{N}_{\alpha,\beta,H}(i)$ 是标准形式的 LFSN 过程。对该模型的系统研究分别由 José R. Gallardo^[5] 和 A. Karasaridis^[6] 等人提出, 随后文^[7,11] 对该模型进行了讨论研究。研究讨论的核心问题之一就是实际网络流量的概率分布问题。为此, 文^[5] 证明了 $|\hat{N}_{\alpha,\beta,H}(i)|$ 仍然具有自相似性, 采用了对称 Alpha-stable 分布取绝对值的方式, 对网络流量进行描述, 以保证模型的实际意义和正确性。文^[6] 假设稳定分布服从完全偏斜的 Alpha-stable 分布, 即 TSA α S (Totally skewed α -stable) 模型, 最大程度地减小负值出现的概率。文^[7] 在文^[6] 的基础上, 放松了对偏斜因子 β 的限制, 以使模型具有最大的灵活性。而文^[11] 为了便于基于模型的分析, 假设稳定分布服从对称 Alpha-stable 分布。

3.2 差分分析方法

通过上述分析可知, 文^[5~7,11] 从各自的研究点出发, 对网络流量模型采用了不同的假设前提, 主要分歧集中在偏斜因子 β 的取值上。这样导致了研究者不得不面对一个棘手的难题: 针对某个特定的实际网络流量, 应该采用何种参数的概率分布才能最有效地反映流量特征。同时, 当前对 Alpha-stable 分布线性分析理论系统研究的成果^[10], 主要针对对称 Alpha-stable 分布。模型的多样性直接导致, 建立在流量模型之上的网络流量预测、网络业务准入机制等方法, 不能利用当前已有的理论和方法。鉴于此, 本文提出了差分分析方法, 从而有效地解决了这些问题。文^[6] 推导了描述基于完全偏斜的 Alpha-stable 分布模型的单位时间流量 PDF 对应的 CF。本文首先给出一般意义下的 Alpha-stable 分布模型的单位时间流量概率分布的特征函数。由 Alpha-stable 分布的特征函数定义和基于 LFSN 过程的网络流量模型的基本形式, 可得 $\hat{N}_{\alpha,\beta,H}(i)$ 对应的随机序列的 CF 为

$$\Phi_N(\theta) = \prod_i \Phi_s(h_d(i) \cdot \theta) = \prod_i \exp\{-|\theta|^{\alpha} |h_d(i)|^{\alpha} (1 - i\beta(\text{sign}\theta) \tan \frac{\pi\alpha}{2})\}$$

即

$$\Phi_N(\theta) = \exp\{-|\theta|^{\alpha} \sum_i |h_d(i)|^{\alpha} (1 - i\beta(\text{sign}\theta) \tan \frac{\pi\alpha}{2})\}$$

从而有

$$\Phi_M(\theta) = \exp\{-|c_1 \cdot \theta|^{\alpha} \sum_i |h_d(i)|^{\alpha} (1 - i\beta(\text{sign}\theta) \tan \frac{\pi\alpha}{2})\}$$

$$\frac{\pi\alpha}{2}) + jc_2\theta\}$$

易知 $\sigma' = c_1 \cdot (\sum_i h_d^{\alpha}(i))^{\alpha}$, $\mu' = c_2$, $\alpha' = \alpha, \beta' = \beta$ 。可见, $\hat{M}(i)$ 服从 $S_{\alpha'}(\sigma', \beta', \mu')$ 。

令序列 $\{x(1), x(2), \dots, x(n)\}$ 服从 $\hat{M}(i)$ 的随机序列, 即 $x(i) \sim S_{\alpha'}(\sigma', \beta', \mu')$, 则有

$\{x(1+k), x(2+k), \dots, x(n+k)\}, x(i+k) \sim S_{\alpha'}(\sigma', \beta', \mu')$ 。根据 Alpha-stable 分布基本性质:

性质 1 若 X_1 和 X_2 是服从 Alpha-stable 分布的独立随机变量, 即 $X_1 \sim S_{\alpha}(\sigma_1, \beta_1, \mu_1); X_2 \sim S_{\alpha}(\sigma_2, \beta_2, \mu_2)$, 则 $X_1 + X_2 \sim S_{\alpha}(\sigma, \beta, \mu)$, 并且 $\beta = \frac{\beta_1\sigma_1^{\alpha} + \beta_2\sigma_2^{\alpha}}{\sigma_1^{\alpha} + \sigma_2^{\alpha}}, \sigma = (\sigma_1^{\alpha} + \sigma_2^{\alpha})^{1/\alpha}, \mu = \mu_1 + \mu_2$ 。

性质 2 若 X 是服从 Alpha-stable 分布的独立随机变量, 即 $X \sim S_{\alpha}(\sigma, \beta, \mu)$, 并且 $c \in \mathcal{R}$, 则

$$\begin{cases} cX \sim S_{\alpha}(c|\sigma, \text{sgn}(c)\beta, c\mu), & \alpha \neq 1 \\ cX \sim S_1(c|\sigma, c\mu - \frac{2}{\pi}c(\ln|c|)\alpha\beta, c\mu), & \alpha = 1 \end{cases}$$

令 $k=1$, 可得

$$x(i+1) - x(i) \sim S_{\alpha}(2^{1/\alpha}\sigma', 0, 0) \quad (2)$$

即网络单位时间内的流量增量服从对称 Alpha-stable 分布。此处采用一般意义上的 LFSN, 能消除概率分布多样性的不利影响, 同时得到流量增量服从对称 Alpha-stable 分布, 可以很好地利用当前基于对称 Alpha-stable 分布的线性分析理论^[10], 对网络流量变化进行定量分析和预测。

3.3 基于差分分析的模型参数估计方法

$c_2 = \mu, H$ 由网络流量的均值和自相似参数得到。 $c_1 = \sigma'$ 通过(2)式得到。此处的实验数据集为差分数据, $S_{\alpha'}(2^{1/\alpha}\sigma', 0, 0)$, 有很强的相关性, Alpha-stable 分布参数估计的一般方法在此处不再适用。log SA α S 方法^[12] 的相关遍历性 (correlation-ergodic), 使得它适合于此处的具有相关性数据序列的参数 α' 和 $\sigma'' = 2^{1/\alpha}\sigma'$ 估计。

偏斜因子 β 决定了模型概率分布函数的偏斜程度, 是参数估计的关键。根据 Alpha-stable 分布的众数公式^[13]:

$$\text{mod } e = \mu + \sigma' * \alpha * \beta * \exp(-b * |\beta|) \quad (3)$$

其中: $a = 1.7665114 + 1.8417675 * \alpha - 2.2954390 * \alpha^2 + 0.4666749 * \alpha^3$

$$b = -0.003142967 + 632.4715 * \alpha * \exp(-7.106035 * \alpha)$$

将 μ, σ, α 带入式(3), 得到偏斜因子 β 的取值的列表。实际估计时可以通过众数值查表得到 β 。

4 仿真结果分析

4.1 Alpha-stable 分布 PDF 分析

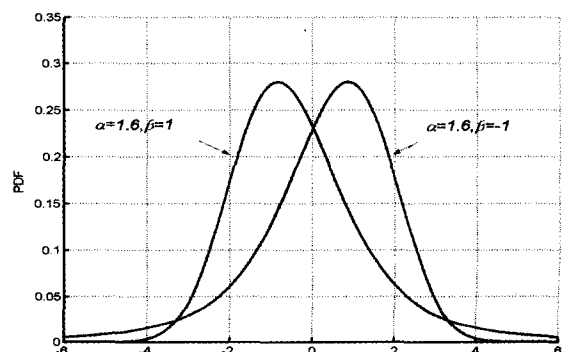


图 1 $\beta = \pm 1$ 的 $X \sim S_{1.6}(1, \beta, 0)$ PDF

图 1,2 描述了 Alpha-stable PDF 随参数 α, β 值变化的情况。图 1 中为 $\alpha = 1.6, \beta$ 分别等于 1 和 -1 时的 Alpha-stable PDF, 可以明显看出偏斜 Alpha-stable PDF 拖尾的不对称性, 其中 $\beta = 1$ 时对应 $TS_{\alpha S}$; 图 2 为 α 值变化时对称 Alpha-stable 分布, $X \sim S_{\alpha}(1, 0, 0)$, PDF 拖尾情况。

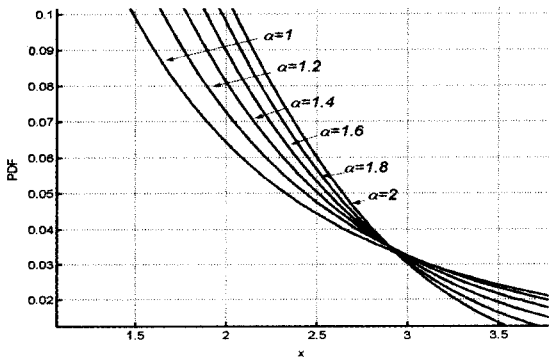


图 2 不同 α 值的 $X \sim S_{\alpha}(1, 0, 0)$ PDF 拖尾

4.2 差分分析参数估计仿真实验

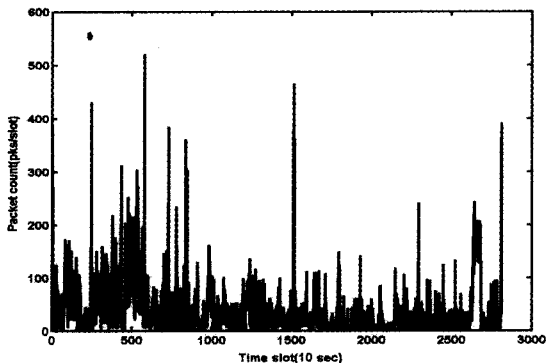


图 3 OctExt 真实流量

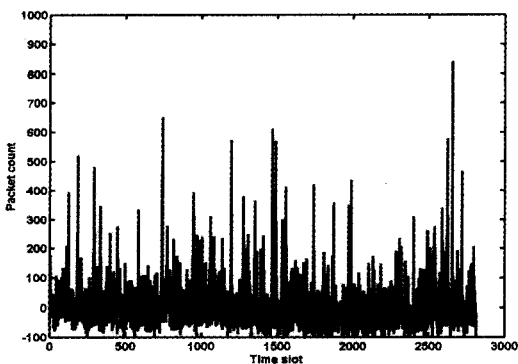


图 4 模拟流量

此处将 Bellcore 采集的实际网络流量 OctExt. TL 与模拟流量进行对比, 然后对其进行拟合分析。Monta-Carlo 实验中采用了基于上述的离散 LFSN 形式生成的 LFSN 随机数, 其中采用 Chambers 方法产生服从 Alpha-stable 分布的随机数。为了精确反映实验结果, 实验采用对经验概率密度函数进行拟合和 QQ 图两种方式对数据进行数据检验。

比较图 3、图 4 可知, 模拟网络流量能较好描述网络流量的突发性。由如图 5 所示的经验概率密度函数拟合图可见, OctExt. TL 数据的分布和本文所得估计模拟流量分布拟合程度较好。同时图 6 中的 QQ 拟合图也进一步证明了分位数拟合的正确性。

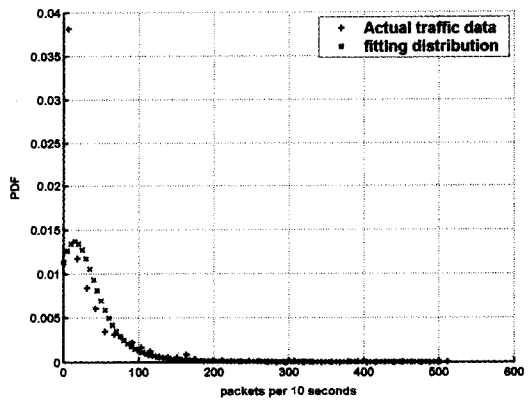


图 5 经验概率密度函数拟合

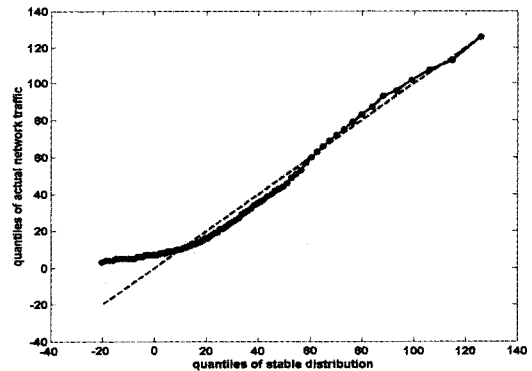


图 6 QQ 拟合

总结 网络流量模型的意义在于为网络资源分配、提高服务质量保证等研究提供理论依据。本文讨论了基于 LFSM 的网络流量模型, 并提出了一种基于差分分析 LFSM 网络流量模型参数的估计新方法。Monta-Carlo 仿真实验得到经验概率密度函数拟合和 QQ 拟合结果, 表明了该方法的有效性和正确性。

参考文献

- 1 Leland W E, Taqu M S, Willinger W, et al. On the Self-similar Nature of Ethernet Traffic (Extended Version). *IEEE/ACM Transactions on Networking*, 1994, 2(1):1~15
- 2 Paxson V, Floyd S. Wide-area traffic: The failure of the Poisson modeling. *IEEE/ACM Transactions on Networking*, 1995, 3(3): 226~244
- 3 Willinger M G. Analysis, modeling and generation of self-similar VBR video traffic. In: *Proc. ACM SIGCOMM'94*, 1994, 269~280
- 4 Taqu M S, Willinger W, Sherman R. Proof of a Fundamental Result in Self-similar Traffic Modeling. *Computer Communication Review*, 1997, 27(2):5~23
- 5 Gallardo J R, Makrakis D, Orozco-Barbosa L. Use of α -stable self-similar stochastic processes for modeling traffic in broadband networks. *Performance Evaluation*, 2000, 40: 71~98
- 6 Karasaridis A, Hatzinakos D. Network heavy traffic modeling using α -stable self-similar processes. *IEEE Transactions on Communications*, 2001, 49(7):1203~1214
- 7 Zhu Guangxi, Ge Xiaohu, Zhu Yaoting. A New Modeling for Network Traffic Based on Alpha-stable Self-similar Processes. In: *The IEEE ICNSP 2003, Nanjin, China, Proceeding*, 2003, 14~17
- 8 Zolotarev V M. *One-dimensional Stable Distributions*. Providence, RI: Amer Math Soc, 1989
- 9 Samorodnitsky G, Taqu M S. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York: Chapman & Hall, 1994
- 10 Nikias C L, Shao M. *Signal processing with Alpha-stable distributions and applications*. New York: Wiley, 1995
- 11 闻勇, 朱光喜. 自相似网络通信量的滑动平均预测. *计算机科学*, 2006, 33(6):32~34
- 12 Lambert P, Lindsey J K. Analysing financial returns using regression models based on non-symmetric stable distributions. *Applied Statistics*, 1999, 48:409~424
- 13 Ma Xinyu. *Applications of the Non-gaussian Stable Process to Communication Engineering*. [Maste]. University of California, 1996