

快速 RMD 法产生的自相似业务流的性质研究^{*}

胡 严, 刘星成, 张光昭
(中山大学电子与通信工程系, 广东 广州 510275)

摘 要: 传统的业务模型大多基于 Poisson 模型或其改进形式, 假定业务突发长度显负指数分布. 近期真实网络流量分析表明很多信息源会产生在长时间尺度下具有自相似特性的信息流. 该性质显著地影响宽带网的流量控制及其排队分析. 合成可控制的自相似业务流是进行仿真的第一步. 实现了一个快速 RMD 算法的自相似流生成器; 估计了产生的自相似流的 Hurst 参数; 估计了两个或多个自相似流叠加后的流的 Hurst 参数; 利用一种打乱算法成功地去掉了自相似流的长相关性.

关键词: 自相似业务; Hurst 参数; RMD

中图分类号: TP393 文献标识码: A 文章编号: 0529-6579 (2001) 03-0029-04

近年来, 对 LAN 和 WAN 业务量的研究表明: 网络业务显现自相似性, 即在所有 (或至少一个较大范围) 时间尺度上表现出统计相似性, 突出表现为突发没有明确的长度, 我们不可能将它平滑掉. 通过对更广泛的不同网络环境业务研究, 例如 Ethernet、ATM、压缩的 VBR 视频和基于 HTTP 的 WWW 业务, 发现自相似性是一种普遍的现象. 传统的网络业务模型没有这种长时相关性, 只是在短时间尺度下才存在突发性, 当业务源数目增加时, 突发性会被吸收, 聚集业务将变得越来越平滑; 而对于自相似模型, 其聚集业务的突发性将更突出而不是减少. 由于自相似问题数学上固有的难度, 目前的研究仍以仿真、测量和数值方法为主. 如何仿真出与实际业务流近似的数据流成为进行网络自相似研究的首要解决的问题. 本文在这方面进行了一定的研究.

1 自相似过程的数学描述

考察一个协方差平稳 (又称为广义平稳) 统计过程 $X = \{x_t; t = 0, 1, 2, 3, \dots\}$, 也就是说, 该过程的数学期望 $\mu = E[x_t]$ 为常数, 方差 $\sigma^2 = E[(x_t - \mu)^2]$ 为有限值, 自相关函数 $r(k) = E[(x_t - \mu)(x_{t+k} - \mu)] / \sigma^2, k = 1, 2, 3, \dots$; 仅与 k 有关. 其中 x_t 可理解为第 t 个单位时间里到达的网络业务实体 (packet, cell 等) 数目. 我们假定 X 的自相关函数有如下形式:

$$r(k) \sim \alpha k^{-\beta}, 0 < \beta < 1, \text{ 当 } k \rightarrow \infty$$

其中, α 为一正常数. 对于每一个 $m = 1, 2, 3, \dots$, 令 $X^{(m)} = \{x_t^{(m)}; t = 0, 1, 2, 3, \dots\}$ 表示一个新的序列, 其中 $x_k^{(m)} = 1/m(x_{km-m+1} + \dots + x_{km}), k = 1, 2, 3, \dots$ 代表长度为 m 的聚集过程. 如是 $X^{(m)}$ 也定义了一个广义平稳随机过程, $r^{(m)}(k)$ 表示 $X^{(m)}$ 的自相关函数. 如果对所有 m , 均有

$$r^{(m)}(k) = r(k) \sim k^{-\beta}, 0 < \beta < 1, \text{ 当 } k \rightarrow \infty$$

我们称 X 为精确二阶自相似过程, 并称 $H = 1 - \beta/2$ 为其自相似参数. 也就是说, 如果聚集过程 $X^{(m)}$ 与过程 X 是不可区分的 (最少是二阶统计特性不可区分), 我们就称 X 是精确自相似过程. 如果

$$r^{(m)}(k) \rightarrow r(k) \sim k^{-\beta}, \text{ 当 } m \rightarrow \infty, k = 0, 1, 2, 3, \dots$$

则称 X 为渐进二阶自相似过程.

自相似参数 H 又称为 Hurst 参数, 是描述自相似特征的唯一参数. 对于 H 有 3 个不同物理意义的取值范围, $0 < H < 1/2$ 表示负相关, $1/2 < H < 1$ 为正相关, $H = 1/2$ 为没有相关性. 实际网络业务量是正相关的, 所以 H 的取值范围是 $(1/2, 1)$, H 越大, 过程自相似程度越高. 自相似过程的最重要特征是: 当 $m \rightarrow \infty$, 其聚集过程 $X^{(m)}$ 的自相关结构并不退化. 传统业务模型则不同, 当 $m \rightarrow \infty$, 聚集过程 $X^{(m)}$ 的自相关结构将退化, 即 $r^{(m)}(k) \rightarrow 0, \text{ 当 } m \rightarrow \infty, k = 1, 2, 3, \dots$.

2 自相似业务的产生方法

分形高斯噪声 (Fractional Gaussian Noise) 经

^{*} 收稿日期: 2000-09-01; 作者简介: 胡严 (1970—), 男, 博士研究生; 通讯联系人: 张光昭.
(C)1994-2019 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

常用在网络仿真中, 网络的业务流可用 FGN 来建模, 它是一个自相似过程, FGN 的增量过程就是分形布朗运动 (Fractional Brownian Motion) .

有几种产生自相似业务流或 FGN 流的方法. 它们是:

- 1) 随机中点置位法 (Random Midpoint Displacement) 法^[1];
- 2) Fourier 变换法^[2];
- 3) Pareto 分布法^[3~5];
- 4) Markov 过程法^[6].

3 对参数 H 的评估

由于自相似过程难于进行排队解析, 因此利用仿真的方法研究自相似业务对高速网络流量控制带来的影响已成为网络设计的一个重要的手段. 要对自相似业务流进行仿真, 首先需要一个好的性能良好的业务流产生器, 而且必须知道该产生器产生的自相似业务流的性质. 已知由产生器产生的或从实际网络中采集的时间序列 X 是自相似的, 那么如何准确地估计该自相似过程的参数 H 呢? 目前常用的估计方法主要有^[7] (由于篇幅的限制, 其他的方法见文献 [7]):

- 1) 方差时间法. 画出 $\lg(\text{Var}(X^m)/\text{Var}(X))$ 和 $\lg(m)$ 的坐标图. 应该得到一条直线, 该直线的斜率为 $-\beta$, β 在 $(0, 1)$ 之间取值, 而 $H = 1 - \beta/2$, 所以我们可以比较简单地估计出 H 值. 它是一个计算量不大, 且精度比较高的方法.

- 2) R/S 分析法. 自相似过程具有 Hurst 效应, 如是可画出 $\lg(E(R(n)/S(n)))$ 和 $\lg(n)$ 的坐标图, 该图应该是以 H 为斜率的一条直线. 它的计算量比较大, 精度不如方差时间法.

4 快速 RMD 法产生的自相似业务流及其业务流的性质

文献 [1] 给出了一种快速 RMD 算法, 该算法首先需要产生一个均值为 0 方差为 1 的正态分布随机数组 G_i , 然后通过求左右两点的算术均值, 迭代地获得中点的数值 $x[n]$, 即

初始值: $x[0] = 0$

初始值: $x[1] = G_1$

叠代 1: $x[1/2] = (x[0] + x[1])/2 + rG_2$

叠代 2: $x[1/4] = (x[0] + x[1/2])/2 + r^2G_3$

叠代 2: $x[3/4] = (x[1/2] + x[1])/2 + r^2G_4$

叠代 3: $x[1/8] = (x[0] + x[1/4])/2 + r^3G_5$

叠代 3: $x[3/8] = (x[1/4] + x[1/2])/2 + r^3G_6$

叠代 3: $x[5/8] = (x[1/2] + x[3/4])/2 + r^3G_7$

叠代 3: $x[7/8] = (x[3/4] + x[1])/2 + r^3G_8$

:

这个算法的优点就是能快速地产生自相似业务流, 而且易于控制 H 参数. 使用中心极限定理^[8] 来获得正态分布随机数组 G_i . 我们用 C++ 编写的能产生不同相关性的自相似业务流的生成器拥有完整的数据产生和处理功能. 向它输入不同的 Hurst 参数 H 和记录 FGN 的数组 $X(n)$, 其中 $n = 1, 2, 3 \cdots N$, 就能产生 Hurst 参数为 H 的自相似流 $X(n)$, 可以把 $X(n)$ 理解为时间间隔 t 内所测量到的字节数. 只有弄清楚了该生成器类所产生的自相似业务流的性质, 才能把它用于以后的网络仿真实验.

对于人工产生的自相似业务流, 我们希望能够控制它的相关性, 这样就需要研究该生成器的输入 H 参数与其生成的流的 H 参数的关系. 用该生成器生成了多组不同的 $N = 8\,388\,608$ 的业务流 $X(n)$, 其中 $n = 1, 2, 3 \cdots N$, 利用方差时间法估计业务流的 H 参数值, 采用最小二乘法拟合, 可得到 $y = \text{slope} \cdot x + \text{shift}$ (例如, $H = 0.80$ 时, $y = -0.4078x - 0.0052$), 如是 $H = 1 + \text{slope}/2$. 每一个数组项 $X(n)$ 代表 10^{-3} s 内到达的字节数, 对业务流进行了 18 种不同尺度的聚集, 取平均到达速率为 150 Mb/s. 实验结果见表 1. 从表 1 可以看出, 用这种方式产生自相似业务流, 完全可以比较精确地控制业务流的 Hurst 系数. 输入生成器的 H 参数在 $(0.5, 0.8]$ 内取值时, H 参数与参数 H 的误差不大于 0.01; 大于 0.80 时, 它们的误差有所增大. 在正相关性的取值范围内, H 参数总是略小于 H 参数. 图 1a 是对生成器合成的具有 5 个不同 Hurst 参数的流使用方差时间法估计 H 参数的结果.

两个渐进自相似业务流叠加而成的合成业务流 Hurst 参数, 理论上是取这两个流的大者. 用自相似流生成器进行了人工合成流的相加实验, 见表 2. 表 2 的第一行是进行相加的 2 个 Hurst 参数, 下面 3 行是对合成出来的业务流 Hurst 参数进行估计的结果, 最后一列是让 $H = 0.60, 0.65, 0.70, 0.75, 0.80$ 的 5 个流相加后的结果, 采用方差时间估计法和最小二乘法拟合. 图 1b 是使用方差时间法估计叠加流的 Hurst 参数的结果.

表 1 参数与生成业务流的参数

Tah 1 The H parameters and the theoretical H parameters

H	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.99
shift	-0.0040	0.0009	0.0040	0.0025	0.0030	-0.0035	-0.0052	0.0070	0.0090	0.0177	0.0241
slope	-0.9984	-0.9049	-0.8104	-0.7125	-0.6144	-0.5098	-0.4078	-0.3212	-0.2262	-0.1510	-0.1068
H	0.5008	0.5476	0.5948	0.6438	0.6928	0.7451	0.7961	0.8394	0.8864	0.9245	0.9466

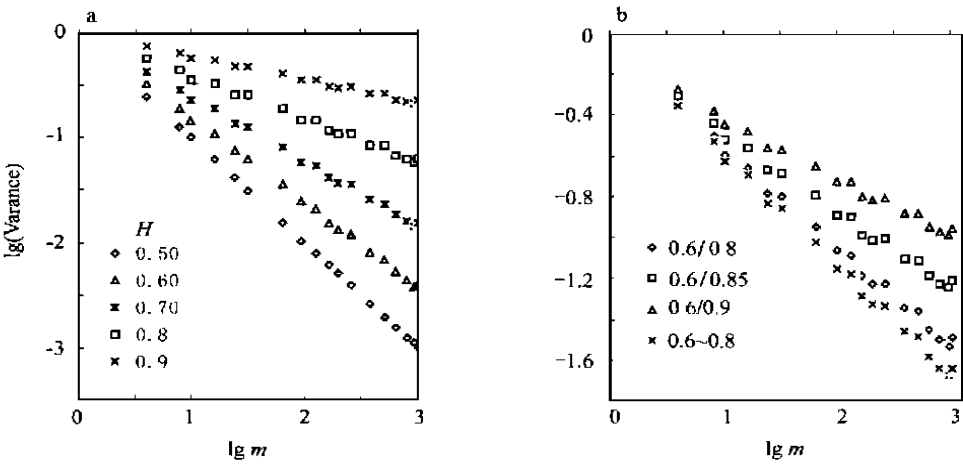


图 1 方差时间法估计 Hurst 参数(a) 和叠加流 Hurst 参数(b)

Fig 1 H parameters(a) and H parameters of the aggregated traffic(b) by var time

表 2 自相似流相加的结果

Tah 2 The results of aggregated self-similar traffic

	0.60/0.80	0.60/0.85	0.60/0.90	0.70/0.80	0.75/0.85	0.80/0.90	0.6~0.8
shift	-0.0808	-0.1115	-0.1364	-0.0262	-0.0243	-0.0145	-0.0555
slope	-0.4857	-0.3814	-0.2864	-0.4789	-0.3791	-0.2980	-0.5415
H	0.7571	0.8093	0.8568	0.7605	0.8104	0.8510	0.7293

由表 2 可以明显地看出：对这种生成器合成的业务流相加而得到的业务流的 Hurst 参数与理论预测有较大的误差，例如，把 $H_1 = 0.6$ 的业务流 X' 和 $H_2 = 0.8$ 的业务流 X'' 相加，即 $X' + X''$ 得到的业务流，这个业务流的 Hurst 参数按理论预测应该为 0.8，但是我们实验的结果是 0.7571。实验所取的两个业务流是各自独立生成，所以可以认为是两个互不相关的过程。出现这种现象的原因可能是：① 我们合成的业务流只是一种近似的渐进自相似过程，这样相加业务流取大 H 参数的条件不满足；② 快速 RMD 算法生成的业务流的 H 参数偏小，用它们相加而得到的流自然也会变小。可见用快速 RMD 算法进行业务流相加来仿真实际网络业务流量时，切不可简单地取大者的 H 参数。

为了研究业务流量之间的相关性对网络性能的影响，可以对实际网络上收取的业务流进行处理，去掉它们之间的相关性，再对它们进行仿真，让得到的结果与原来业务流的结果对比，就能定量地测出相关性对网络性能的影响^[9]。用生成器产生没有相关性的业务流有两种方法，第一就是直接向生成器输入 $H=0.5$ ，那么得到的就是一个近似无相关性的业务流；第二就是把生成器得到的业务流进行打乱，即随机地排列 $X(n)$ 的先后次序，这样能完全地去掉业务流之间的相关性。在我们的软件包内提供了这样的打乱算法（即，一种交织器算法，已在 [10] 中进行了讨论），如果使用 32 位的计算机，它能处理 $N=2^{12}$ 长的业务流。由表 1 可以看到，生成器合成的无相关性业务流（即， $H=0.5$ ），有一定的正相关性

(H 略大于 0.5). 而第二种方法可获得比较好的无相关性业务流. 用 $N = 2^{15}$ 长的自相似流进行了实验, 去掉统计误差, 把这些流打乱后, 它们的 Hurst 参数的均值为 0.5, 这种流可以完全满足仿真实验的需要.

5 总结和结论

用 C++ 实现了一个快速 RMD 算法的自相似流生成器, 而且编写了多种估计 Hurst 参数方法的功能模块. 利用它们, 我们研究了快速 RMD 算法产生的自相似流的性质: ①该算法产生的自相似流的 H 参数在 (0.5, 0.8] 内取值时, H 参数与 H 参数的误差不大于 0.01, 大于 0.80 时, 它们的误差有所增大. 在正相关性的取值范围内, H 参数总是略小于 H 参数; ②两个或多个流叠加后, 其 Hurst 参数与理论值有较大的误差, 给出了初步的解释, 但还需更进一步的理论研究; ③利用一种交织器算法成功地去掉了自相似流的长相关性.

参考文献:

- [1] DROZ P, BOUDECE J L. A high-speed self-similar ATM VBR traffic generator[J]. IEEE Global Telecommunication Conference, 1996(11): <http://www.zurich.ibm.com/~dro/globecom96.ps.z>.
- [2] PAXSON V. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic[J]. Computer Communications Review, 1997(10): 5—18.
- [3] CROVELLA M K, BESTAVROS A. Self-similarity in World Wide Web traffic: evidence and possible causes[J]. IEEE/ACM Transactions on Networking, 1997, 5(6): 835—846.
- [4] PARK K, KIM G, CROVELLA M K. On the relationship between file sizes, transport protocols, and self-similar network traffic[J]. IEEE International Conference on Network Protocols, 1996: 171—180.
- [5] TUAN T, PARK K. Multiple time scale redundancy control for QoS-sensitive transport of real-time traffic[J]. IEEE INFOCOM, 2000, Tel-Aviv, Israel. <http://www.ieee-infocom.org/2000/papers/725.ps>.
- [6] NORROS I. On the use of fractional Brownian motion in the theory of connectionless networks[J]. IEEE JSAC, 1995, 13(6): 953—962.
- [7] TAQQU M S, TEVEROVSKY V, WILLINGER W. Estimators for long range dependence: an empirical study[J]. Fractals, 1995, 3(4): 785—788.
- [8] 茆诗松, 王静龙, 濮晓龙. 高等数理统计[M]. 北京: 高等教育出版社, 施普林格出版社, 2000. 400—401.
- [9] MORIN P R, NEILSON J. The impact of self-similarity on network performance analysis[J]. Carleton University, Computer Science Department, 1995. <http://cgm.cs.mcgill.ca/~morin/reports/honours.ps>.
- [10] 刘星成, 张光昭. 一类新的 turbo 码交织器设计[J]. 中山大学学报(自然科学版), 2000, 39(6): 39—43.

Study of the Characters of Self-Similar Traffic Stream Generated by High-Speed RMD Method

HU Yan, LIU Xing cheng, ZHANG Guang zhao

(Department of Electronics and Communication Engineering, Zhongshan University, Guangzhou 510275, China)

Abstract: The conventional models are mostly based on Poisson models or their improved versions. It is assumed that the distribution of burst length is exponential. Recent traffic analyses of a lot of networks have shown that many traffic sources produce traffic streams that are self-similar over several time scales. This character has severe impact on flow control and queuing analysis in broadband networks. A self-similar traffic stream generator using a high-speed RMD algorithm has been implemented. The Hurst parameters are estimated for those self-similar traffic streams from the generator and several superposed self-similar traffic streams. The autocorrelation structure of the self-similar traffic is destroyed by shuffling algorithm.

Keywords: self-similar traffic; Hurst parameter; RMD