

# 基于高速网络控制的阈交方法

刘嘉<sup>1</sup>, 柳湘月<sup>1</sup>, 王 旭<sup>1</sup>, 舒炎泰<sup>2</sup>

(1. 天津大学理学院, 南开大学天津大学刘徽应用数学中心, 天津 300072; 2. 天津大学电子信息工程学院, 天津 300072)

**摘 要:** 在高速网络控制与带宽分配研究中, 网络业务量预报是一个重要的问题. 本文给出了一个新方法, 通过对高速网络业务超阈值数据采用方差分析的方法构造了一个阈值的能量函数, 为阈值的选取提供了一定的依据. 最后运用阈交理论, 通过对网络业务到达数据对阈值穿越强度的计算来考察已选阈值的有效性.

**关键词:** 网络业务; 阈交问题; 阈值; 穿越强度

**中图分类号:** TN91312 **文献标识码:** A **文章编号:** 03722112 (2003) 12170204

## The Threshold Crossing Method on High2Speed Network Control

LIU Jia2kun<sup>1</sup>, LIU Xiang2yue<sup>1</sup>, WANG Xu<sup>1</sup>, SHU Yan2tai<sup>2</sup>

(1. School of Science, Tianjin Univ. LiuHui Center for Appliad Mathematics, Nankai Univ. & Tianjin Univ, Tianjin 300072, China ;

2. School of Electronic Information Engineering, Tianjin Univ, Tianjin 300072, China )

**Abstract:** Network traffic prediction is important for network control and allocation. This paper presents a new method. The energy function of threshold value  $u$  to be constructed by the method of variance analysis for high2speed network traffic, presents bases for selecting threshold value. At last, according to threshold crossing theory, we examine the validity of the solution through calculating the crossing intensity.

**Key words:** network traffic; threshold crossing problem; threshold value; crossing intensity

### 1 引言

随着计算机通讯业务的日益发展, 提供业务服务质量保证(QoS)成为一个重要的研究方向, 而网络拥塞检测和控制是保证QoS的关键技术. 在网络业务性能评价的研究过程中, 网络业务模型在很长一段时间都采用传统的话务模型或其改进形式. 在最初研究ATM网络时, 采用的也是Poisson过程, 随着研究的深入逐步引入了一些较为复杂的随机模型, 如fluid flow模型、packetrain模型、Markov2modulated Poisson过程、批到达Markov过程等. 这些模型的共同特点是所描述的业务序列具有短时相关性(short range dependence), 当时间标度增加时, 统计上单位时间内得到的数据包数将趋于白噪声. 十年前, 由美国贝尔实验室和Boston大学研究人员对以太网业务和VBR视频业务的研究结果表明: 实际网络业务普遍存在统计上的自相似性, 具有长相关性(long range dependence)<sup>[1]</sup>. 近年来, 各国研究人员对其他网络业务进行了测量和分析, 均发现了网络业务具有自相似性. 由于传统业务模型描述能力的欠缺, 它们不能描述业务的自相似性. 当业务源增加时, 按照传统模型得到的结论是聚集业务越来越光滑; 而对于自相似业务其聚集业务的突发性将更突出而不是减少. 最近的研究结果表明,

较好的模型如分数布朗运动, FARIMA过程等模型在网络建模, 预报和准入控制方面提供了一个较好的方法<sup>[2,3]</sup>. 但由于其业务模型相当复杂. 这使得基于业务模型预报的网络控制由于计算的繁杂而相当困难<sup>[4]</sup>. Duffield提出了在长相关业务FBM(分数布朗运动)下, 应用大偏差理论(Large Deviation)分析队列长度尾分布性质的方法<sup>[5]</sup>. 张连芳等人证明了大偏差技术也适用于FARIMA业务尾分布的渐近分析, 并给出大偏差在分析队列系统时的重要结论<sup>[6]</sup>. 但由于数学解析的复杂性, 实用的参数估计方法很难得到.

由于在网络性能评价研究中, 通过优化带宽分配解决拥塞和溢出等问题时, 我们更关心的是网络业务超出某一阈值的情况, 而对阈值以下的业务显然与要解决的问题关系不大. 我们提出采取阈交的方法可以不考虑业务的概率分布, 避开业务模型数学处理的困难, 是一种可行的新方法. 作为一种阈交问题, 有关阈值选取的研究是重要的. 通过对阈值选取的研究, 我们就可以在一定范围内, 近似估计出网络业务到达数量这一随机数据超阈值部分在整体网络随机数据中所占的比例, 及其对网络拥塞的影响, 从而使网络业务在保证服务质量的同时尽可能的减小缓存区, 以达到节省空间的目的, 实现合理的资源配置.

收稿日期: 20020427; 修回日期: 20030319

基金项目: 国家自然科学基金(重大No. 90104015); 天津市自然科学基金(No. 023601111); IBM SUR; 南开大学天津大学刘徽应用数学中心(No. T08)

本文提出采用方差分析的方法构造一个阈值的能量函数,并研究了超阈值能量函数的性质.并以文[1]pAugIL#1s(3千多数据)和pAugIL#100(3万多数据)为例进行实证分析,进而得出规律,为阈值的选取提供一定的依据.

## 2 阈值的选取

### 2.1 阈值的初始选取范围

由于高速网络业务具有分形的特征,阈交问题主要研究突发事件,也即极值问题,故确定阈值的选取范围是很有必要的,也是很有实际意义的.

若 $\{X_n, n=1, 2, \dots, N\}$ 是一平稳随机过程,设其均值、方差分别为

$$m = E[X_n] = \frac{1}{N} \sum_{i=1}^N X_i$$

$$s^2 = D[X_n] = \frac{1}{N} \sum_{j=1}^N (X_j - E[X_j])^2$$

其中 $s$ 为标准差.又设 $M_k$ 为 $\{X_n\}$ 中的最大值.由于 $s$ 表示数据在整体对均值 $m$ 的偏差程度,因此作为对突发事件的研究,在 $(m+s, M_k)$ 中选取阈值,比在 $(m, M_k)$ 中选取更实际有效.

在以下的分析中,都在 $(m+s, M_k)$ 中选取初始阈值.

### 2.1.2 对不同阈值的整体方差分析

若 $\{X_t, t=1, 2, \dots, T\}$ 为一网络业务到达数据,那么 $\{X_t\}$ 为一平稳随机过程.定义阈值 $u$ 的能量函数 $f(u)$ 为

$$f(u) = \frac{1}{T} \sum_{i=1}^T (X_i - u)^2 \tag{1}$$

那么 $f(u)$ 刻画了业务到达数据在整体上对阈值变量 $u$ 的偏差.由于

$$f'(u) = \frac{2}{T} \sum_{i=1}^T (u - X_i) = 2u - \frac{2}{T} \sum_{i=1}^T X_i$$

为一条直线,可知 $f(u)$ 的变化趋势是非线性的.容易看出 $f(u)$ 具有平滑上升的曲线趋势,显然网络业务到达的实际状况与理论推导结果是吻合的.由于网络业务具有自相似性,那么以上两组数据的 $f(u)$ 值都具有平滑递增的趋势.由于它的变动是不剧烈的,因此在以下的讨论中,不将其作为主要的参考对象.

### 2.1.3 对不同阈值的超阈值数据方差分析

若 $\{X_t, t=1, 2, \dots, T\}$ 为一网络业务到达数据,它是平稳随机过程.对这一随机过程作如下假设:

(1)  $\{Y_t(u), t=1, 2, \dots, S\}$ 为 $\{X_t\}$ 中超过阈值 $u$ 的那部分数据的集合,显然它随 $u$ (由实际意义知必为整数)的增大,它是单调不增函数,即有

$$\{Y_t(u)\} = \{Y_t(u+1)\} = \{Y_t(u+2)\} = \dots;$$

(2)  $N_X(u)$ 为 $\{X_t\}$ 中超过阈值 $u$ 的个数,即集合 $\{Y_t(u)\}$ 的项数,那么 $N_X(u)$ 满足

$$N_X(u) \setminus N_X(u+1) \setminus N_X(u+2) \setminus \dots;$$

(3) 令 $g(u)$ 表示随 $u$ 的变化 $\{X_t\}$ 中超阈值部分对 $u$ 的偏差函数,那么如式(1)有

$$g(u) = \frac{1}{N_X(u)} \sum_{i=1}^{N_X(u)} (X_i(u) - u)^2 \tag{2}$$

$g(u)$ 的函数性质由于超阈值数据间隔的不确定性,采用对数据进行图形分析的方法来初步确定其性质.应用软件包 $\Sigma 2Plus$ 进行了相关的计算与绘图<sup>[7]</sup>.

首先对pAugIL#1s中的前1000个数据及全部数据分别作式(2)的计算.通过绘图发现,无论是小样本数据还是大样本数据,其函数都明显的呈现先迅速下降,然后在某一区域明显提升,最后下降至零,这一剧烈的变动过程见图1.

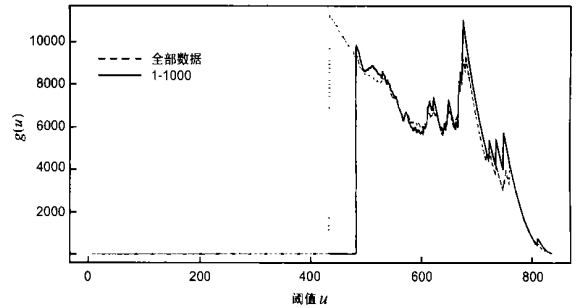


图1 超阈值部分对 $u$ 的偏差函数

显然,两条线极其相似,这正说明了网络业务具有自相似性.为进一步确定函数的独特性质,又对数据pAugIL#100的前1000项、1000~2000项及2000~3000项全部数据分别进行了同样的处理,发现它们几乎都具有与图1完全相似的性质,将其中前三组数据的函数对比着画于同一幅图中(图2).

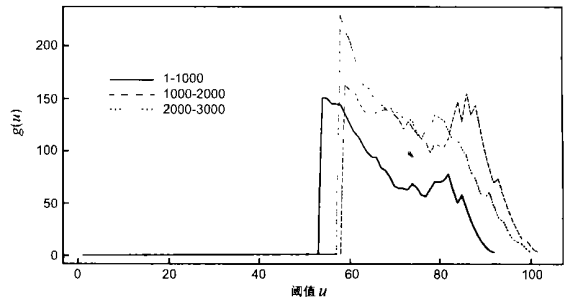


图2 三组数据的超阈值部分对 $u$ 的偏差函数

经过以上对两组数据的图形分析,可以近似得出: $g(u)$ 函数具有先下降,然后在某一区间明显向上提升、最后下降至零的特性.

### 2.1.4 运用 $g(u)$ 函数及 $N_X(u)$ 函数来确定阈值

由于 $g(u)$ 对比于 $f(u)$ 具有较特殊的性质,因此认为 $g(u)$ 函数的特性对阈值的选取一定具有某种意义.通过进一步把 $g(u)$ 和 $N_X(u)$ 进行对比,又发现, $g(u)$ 函数在一定区间上的向上提升大都出现在 $N_X(u)$ 相对较小且下降速度减缓处,见图3.

通过对大量的类似于图3形式的图的观察,我们认为 $g(u)$ 在向上提升时 $u$ 值的统计意义在于:若 $u$ 值继续扩大,反而会在一定范围内使得超阈值部分对阈值 $u$ 的偏差程度加大.结合网络业务到达数据的实际意义,那就意味着此时 $u$ 值的加大是没有效率的,过大的阈值对网络的控制是一种资源浪费,因此不宜再加加大.从上面的讨论就可以初步确定,较好的值应取在提升过程之前.从而进一步缩小了阈值的选取范围.

$g(u)$ 在明显提升过程之前大体上为一个下降过程. 虽然存在一定范围内的上下波动, 但这种波动相比较于后面的提升过程是不剧烈的, 因此可将其视为下降过程的一部分来考虑. 而下降和明显提升这两个过程的拐点, 我们不妨近似取作它们中  $g(u)$  值最小处. 通过观察发现: 在这个下降过程中先是急速下降, 这意味着  $u$  值加大一点儿, 超阈值部分对  $u$  的偏差就发生较大变化, 那么就必然要求  $u$  值继续加大; 而后是一个缓慢下降过程, 考虑到网络的实际应用, 为更有效防止拥塞的出现, 希望  $u$  值稍大一些, 因此  $u$  值在这一阶段仍可以在某种程度上继续加大.

再观察  $N_X(u)$ , 在大多数图形中也都是先快速下降而后减缓, 故结合  $N_X(u)$  和  $g(u)$  的意义和函数图, 就可以得出:  $u$  值应取在  $g(u)$  和  $N_X(u)$  都开始下降缓慢之后, 并在  $g(u)$  明

显提升之前. 这样就既可以避免  $u$  值的无谓扩大, 又可以充分保证超阈值部分在整个网络数据中所占的比例, 也就是使绝大多数情况下网络都畅通.

就 pAugTL# 100 这一组数据而言, 分别对前 1000 数据、1000~ 2000 及 2000~ 3000 数据, 通过以上方法进行阈值小区间(不妨提供围绕拐点的可供选择的 5 个  $u$  值)选取. 可以得出在三种情况下的再次缩小范围后的  $u$  值区间分别为 [75, 79], [76, 80], [74, 78], 它们基本上是一致的, 也就是说这种取值方法是有实际意义的. 由于网络业务具有自相似性, 故综合以上三组数据便可近似得到 pAugTL# 100 这组数据一个测度较小的阈值区间 [75, 79]. 虽然这一结果仍是一区间, 但对比于阈值的初始选取范围  $(m + s, M_k)$  (对于 pAugTL# 100 这组数据初始范围为 (50, 157)) 可以说已经是相当精确的了<sup>[4]</sup>.

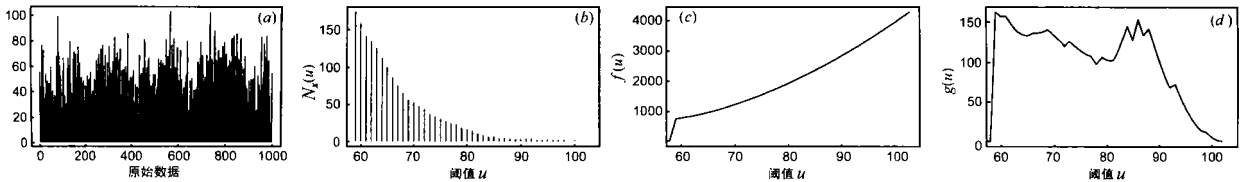


图 3  $g(u)$  和  $N_X(u)$  的对比

### 3 阈值的穿越强度

以上基本上完成了对阈值的选取, 下面采用实际网络业务到达对阈值的穿越强度这一参数来考察已选阈值的有效性. 并通过阈交理论推导出穿越强度公式.

设在网络业务中网络业务到达数量  $\{X(t), t \in T\}$  是一个随机过程, 记  $N(u; t_1, t_2)$  为在某一时间间隔  $[t_1, t_2] \in T$  中,  $X(t)$  与直线  $x = u$  的交点的个数. 由于  $X(t)$  的随机性, 对固定的  $u, t_1, t_2$ ,  $N(u; t_1, t_2)$  是一个随机变量, 那么  $N(u; t_1, t_2)$  为业务到达量穿越阈值  $u$  的次数. 设  $X(t)$  是均方可微的正态过程. 令

$$Y(t) = h[X(t) - u], \quad t \in T \quad (3)$$

其中  $h(x)$  为单位阶跃函数, 即  $h(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$ .

对  $Y(t)$  求导数, 得

$$\dot{Y}(t) = \dot{X}(t) D[X(t) - u] \quad (4)$$

这里  $D(x)$  是 Dirac 的 D 函数, 于是

$$N(u; t_1, t_2) = \int_{t_1}^{t_2} |\dot{Y}(t)| dt = \int_{t_1}^{t_2} |\dot{X}(t)| D[X(t) - u] dt$$

对上式两边取数学期望, 则有

$$\begin{aligned} E[N(u; t_1, t_2)] &= \int_{t_1}^{t_2} E[|\dot{X}(t)| D[X(t) - u]] dt \\ &= \int_{t_1}^{t_2} \int_{-\infty}^{\infty} |\dot{x}| f(t, u; t, \dot{x}) dx dt \end{aligned} \quad (5)$$

其中  $f(t, x; t, \dot{x})$  是  $(X(t), \dot{X}(t))$  的联合密度.  $X(t)$  在  $t$  时与  $u$  相交的平均强度为

$$C(t, u) = \int_{-\infty}^{\infty} |\dot{x}| f(t, u; t, \dot{x}) dx \quad (6)$$

特别在  $\{X(t), t \in T\}$  是平稳时,  $C(t, u)$  与  $t$  无关.

如果进一步再假定  $X(t)$  是平稳的, 则由  $X(t)$  均方可导, 因而  $B(S)$  二次可导<sup>[8]</sup>. 对上式求导可得  $B_c(S) = -B_c(-S)$ . 令  $S = 0$ , 则有  $E[X(t) \dot{X}(t)] = \frac{5}{5t} \#(t, t) = B_c(0) = 0$ . 由于  $X(t)$  是正态的, 所以  $X(t)$  与  $\dot{X}(t)$  独立. 于是  $X(t)$  与  $\dot{X}(t)$  的联合分布密度是

$$f(x, \dot{x}) = \frac{1}{2\pi R_x R_{\dot{x}}} \exp\left\{-\frac{1}{2}\left(\frac{x^2}{R_x^2} + \frac{\dot{x}^2}{R_{\dot{x}}^2}\right)\right\} \quad (7)$$

其中  $R_x^2 = \#(t, t) = B(0)$ ,  $R_{\dot{x}}^2 = \frac{5^2}{5^2 t} \#(t, t) \Big|_{t=s=t} = -B_d(0)$ , 代入式(6)可得

$$\begin{aligned} C(u) &= \int_{-\infty}^{\infty} |\dot{x}| \frac{1}{2\pi R_x R_{\dot{x}}} \exp\left\{-\frac{1}{2}\left(\frac{u^2}{R_x^2} + \frac{\dot{x}^2}{R_{\dot{x}}^2}\right)\right\} dx \\ &= \frac{1}{2\pi R_x R_{\dot{x}}} \exp\left\{-\frac{u^2}{2R_x^2}\right\} \int_{-\infty}^{\infty} |\dot{x}| \exp\left\{-\frac{\dot{x}^2}{2R_{\dot{x}}^2}\right\} dx \\ &= \frac{R_{\dot{x}}}{\pi R_x} \exp\left\{-\frac{u^2}{2R_x^2}\right\} \end{aligned} \quad (8)$$

这样就得到了穿越阈值的强度.

### 4 网络业务对阈值穿越强度的实证分析

我们仍以 pAugTL# 100 为例, 计算出对应于阈值  $u$  的穿越强度  $C(u)$ . 在计算过程中, 将数据离散化, 推导过程应用随机分析要求的条件是满足的. 仍用 S2Plus 计算.

通过分别对 pAugTL# 100 的前 1000 项、1000~ 2000 项及所有数据的统计计算, 得到如下表中一组  $R_x^2$  和  $R_{\dot{x}}^2$  的值.

	1~ 1000	1000~ 2000	All of pAugTL# 100
$R_x^2$	306.2292	347.507	266.4858
$R_{\dot{x}}^2$	250.285	247.2302	177.1396

将各组值代入穿越强度式(8)中,可得如图4所示的曲线.

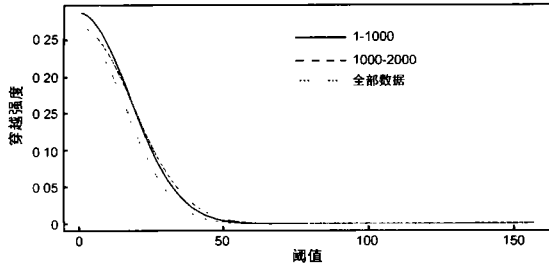


图4 pAugPI#100 穿越阈值强度

可见,三组数据的穿越强度图极其相似,这又从另一个侧面证明了网络业务的自相似性.这里取第一组值即由前1000个数据算出的  $R_x^2$ ,  $R_y^2$  值,代入公式中便可得出数据 pAugPI#100 的穿越强度函数为:

$$C(x_0) = 0.2877692 \exp\{-0.001632764u^2\}$$

对于在上节中求出的阈值,通过对穿越强度的积分进行穿越次数的计算可得,

$$N(u; t_1, t_2) = \int_{t_1}^{t_2} C(t) dt = \int_{t_1}^{t_2} 0.2877692 \exp\{-0.001632764t^2\} dt$$

结合以上所选的阈值区间[75, 79],可以看出在此区间上网络数据对阈值的穿越强度是非常小的,换言之,通过计算证明了所选区间的有效性.避免了无谓增大阈值造成资源的浪费,为优化带宽分配提供了依据.

## 5 结论

在高速网络拥塞控制研究中,由于网络业务模型的复杂性,为溢出概率的计算带来极大的困难,本文给出的阈交方法避开业务模型底分布数学解析的难点,通过对  $f(u)$  及  $g(u)$  性质的讨论,得到了网络业务到达超阈值数据方差的一些性质,并运用这些性质对阈值进行了选取.最后运用阈交强度对选取的阈值进行检验.证明了方法的有效性.应用阈交方法的结果可使网络业务在保证服务质量的同时优化带宽分配,尽可能的减少缓存区,以达到节省空间的目的,实现合理的资源配置.为有关高速网络的准入控制提供一个有价值的依据.本文采用的数据是最具代表性的长相关自相似网络数据,由于提出的方法避开业务数据底分布的复杂性,仅考虑超阈值部分,显然具有一定的普遍性.

## 参考文献:

- [1] W Leland, et al. On the self-similar nature of ethernet traffic (extended version)[J]. IEEE/ACM Transactions on Network, 1994, 2(1): 1-15.
- [2] J K Liu, Y T Shu, L F Zhang, F Xue. Traffic modeling based on FARMA models[A]. IEEE CCECE Piscataway[C]. NJ USA, 1999, 5. 162-167.
- [3] W Willinger, et al. Self-similarity in high-speed packed traffic: Analysis and modeling of ethernet traffic measurements[J]. Statistical Science, 1995, 10(1): 67-85.
- [4] Z G Jin, Y T Shu, J K Liu, Oliver Yang. Prediction-based network bandwidth control[A]. IEEE CCECE, Halifax[C]. Nova Scotia, Canada, 2000, 5. 7-10.
- [5] N G Duffield. Large deviations and overflow probabilities for the general single-server queue, with applications[J]. Math. Proc. Cambridge Phil. Soc. 1995. 363-375.
- [6] 张连芳,等. 自相似业务模型下的队列分析))) 大偏差技术[J]. 通信学报, 1998, 20(4): 23-28.
- [7] W N Venables, B D Ripley. Modern Applied Statistics With SPLUS[M]. 2nd ed. New York, Springer-Verlag 1997. Chap. 15.
- [8] I I Gihman, A V Skorohod. The Theory of Stochastic Processes[M]. Berlin, Springer-Verlag, 1983. Chap. 4.

## 作者简介:



**刘嘉** 男, 1943 年生于天津市, 教授, 主要研究方向为应用随机过程及计算机网络业务性能评价.



**柳湘月** 女, 1978 年生于吉林省敦化市, 硕士研究生, 研究方向为随机过程及其应用.