存储系统负载自相似性研究综述

邹 强 程 强

(西南大学计算机与信息科学学院 重庆 400715)

摘 要 I/O突发是造成I/O瓶颈的一个主要原因,研究I/O负载中普遍存在的突发性并对负载进行精确合成,对存储系统设计及其性能评价具有重要意义。对实际I/O负载的研究表明,传统的泊松假定难以准确地描述长时间范围内的I/O突发行为。研究发现,I/O突发在不同时间尺度下具有相似性,即I/O负载具有自相似性,因此,自相似模型被用来刻画I/O负载中的长相关性。针对I/O负载自相似参数估计,总结了各种常用的时域和频域估值方法。着重对已有的I/O负载合成模型进行了剖析,讨论了各种自相似模型、多分形模型以及alpha稳定模型的特点。探讨了有待解决的开放性问题,并对I/O负载自相似性研究的发展趋势进行了展望。上述工作将对存储负载的自相似性研究提供有益参考。

关键词 存储系统,I/O负载,自相似性

中图法分类号 TP302 **文献标识码** A

Survey of Studies on Self-similarity in Storage System Workload

ZOU Qiang CHENG Qiang

(School of Computer and Information Science, Southwest University, Chongqing 400715, China)

Abstract I/O bursty is one of the main reasons causing I/O bottleneck, so, it is significant for designing storage system and evaluating system performance to study and accurately synthesize the ubiquitous bursty in I/O workload. Research results show that the traditional poisson assumption is difficult to describe the I/O-burstiness behavior well at the long-term time scales, and I/O bursty exhibits the similarity at different time scales, i. e. , self-similarity. So, self-similar models are used to characterize the long-range dependence in I/O workloads. Aimed at the Hurst parameter estimate, this paper summarized the time-domain and frequency-domain estimators usually used to estimate the degree of self-similarity in storage workloads. After that, some existing models synthesizing I/O workloads were examined, thereinto, the characteristics of self-similar, multi-fractal and alpha-stable models were discussed. After summarizing the unresolved problems, this paper explored the future trend of the study on self-similarity in I/O workloads. The above work will provide a valuable reference for pushing the research on self-similarity in storage workloads.

Keywords Storage system, I/O workload, Self-similarity

1 引言

I/O负载特征研究对改善存储子系统性能至关重要。无论是检验存储系统设计的效果,还是比较不同系统设计的性能和功效,都需要对I/O负载特征进行深入的理解和分析。

一方面,处理器的性能近年来正以每年60%的比例按摩尔定律递增,而由于机械延迟的限制,磁盘访问时间等性能每年的改进比却小于10%。随着处理器与磁盘存储之间性能差距的持续扩大,对物理数据流特性进行研究并采取相应对策显得日益重要。

另一方面,伴随着市场需求,最近的存储技术革新如云存储、异构存储管理方法、自优化智能存储技术等正风靡于存储行业。首先,日益处在网络环境下的存储设备,能同时被多个计算机直接共享和访问[1]。其次,存储技术正在发生革命性的变化,催生了新的网络存储技术,如主动存储等[2]。同时,存储系统中可用处理能力的迅速提高,使得建立能根据负载情况进行动态自我优化的智能存储系统成为可能[3,4]。另外,对分布在一个系统内的存储资源,通过各种存储工具采取统一管理的方式已经越来越流行[5]。而这些先进的存储技术要想达到预期效果,尚需要针对真实负载的数据特征进行集中考察,并以此为依据进行性能优化和指导存储系统设计[6]。

在存储负载特征研究和负载合成方面,Ganger通过对块一级I/O负载的研究,做了一系列开创性的工作[7]:分别对访问模式(access pattern)和到达模式(arrival pattern)中的突发性(burstiness)进行分析后发现,在存储负载中通常将访问到达简单假定为泊松到达是与实际情况不相匹配的,"I/O负载自相似性"这一概念被首次提及。

在数据流量中发现自相似性始于对网络流量特征的研

到稿日期:2012-09-05 返修日期:2012-11-08 本文受国家科技支撑计划资助项目(2012BAD35B08),中央高校基本科研业务费专项资金资助项目(XDJK2012A006),重庆市自然科学基金项目(2011BB2008),西南大学博士基金资助项目(SWU111015)资助。

邹 强(1979—),男,博士,副教授,CCF会员,主要研究方向为存储系统负载特征与性能评价,E-mail:qzou@swu.edu.cn;程 强(1971—),男,博士,教授,主要研究方向为高性能计算。

究。在1993年，Leland等人[8]对局域网的网络流量测试数据进行了严格的统计分析，发现网络业务流的突发性在各个时间尺度上都广泛存在，并不像泊松模型描述的那样——网络流量变化的剧烈程度随着时间尺度变大而逐渐被平滑掉，这是自相似现象的一个显著标志。继Leland等人发现在局域网网络流量中存在自相似性以后，学术界广泛研究了实测到的各种网络业务[9-11]，如广域网业务[9]、VBR视频业务[12]、WWW业务[13]等，发现这些网络业务过程中存在的突发行为均呈现出自相似性，或称为长相关性(Long-Range Denpendence，LRD)。

在存储领域，对I/O负载突发行为的研究受到存储研究者的广泛关注和重视。国外的多项研究结果表明，许多实际I/O负载中确实存在自相似性[14-18]。虽然有多种途径可以评估I/O负载的自相似性，且生成自相似I/O负载的模型多种多样，但I/O负载的自相似性研究在某些方面仍需完善。研究发现，目前还没有哪一种自相似性评估方法能给出权威结论，自相似性的评估尚需要深入研究；已有的自相似负载模型通常都缺乏明确的解析表达式，且在模型选取方面，应结合I/O负载的自身特点；采用负载模型合成自相似I/O负载，关键是精确描述自相似负载中的I/O突发，如何建立更精确的负载模型需要进一步研究；将I/O负载的自相似性研究服务于系统设计优化、对系统进行无偏性能评价、Cache算法设计、存储系统的QoS保证等课题，也是尚未解决的难题。

在I/O负载自相似研究综述方面，虽然李明强等人[19]做过出色的工作，但文献[19]中并未介绍近几年最新的研究成果。本文不仅涵盖最新的研究成果，还尝试以不同的内容组织形式对存储系统负载自相似性的研究成果进行介绍。

2 自相似随机过程

虽然对自相似随机过程进行定义的文献多种多样，且一般分为连续时间随机过程和离散时间随机过程，但对数据流量的分析通常被严格地限制为离散时间的二阶或渐近二阶自相似过程。

简单地说，对于变化的时间尺度或空间尺度，概率分布具有不可变性的随机过程称为自相似随机过程。由Leland[8]、Beran[11]等对自相似随机过程的定义可知：考察一个协方差平稳的随机过程 $X=\{X_t\}=(X_1,X_2,\cdots)$，设 X_t 具有恒定均值 $\mu(\mu=E[X_t],X_t=x_t+\mu)$ 和有限方差 $\sigma^2=\mathrm{var}\ X=\mathrm{var}\ x=E[x_t^2](x=\{x_t\})$，且其自相关函数为 $r(k)=E[x_tx_{t+k}]/E[x_t^2](k\geqslant 0)$，其中，$X_t$ 为第 t 个单位时间内到达的访问请求数目，如果 $r(k)$ 具有式(1)所示形式，则称自相关函数满足上述条件的随机过程为自相似过程。

$$r(k)\sim ck^{-\beta} \tag{1}$$

式中，$k\to\infty$，$0<\beta<10$。

若 $X_t^{(m)}=\frac{1}{m}\sum_{i=0}^{m-1}X_{m-i}$，则称 $X^{(m)}=(X_t^{(m)}:t=1,2,3,\cdots)$ 为 $X=(X_t:t=1,2,3,\cdots)$ 的 m 阶聚集过程，其自相关函数记为 $r^{(m)}(k)$。对于协方差平稳的离散随机过程 X_t，如果其 m 阶聚集过程 $X_t^{(m)}$ 具有与原过程 X_t 相同的自相关函数结构，即 $r^{(m)}(k)=r(k)$ 恒成立，则称 X_t 为严格二阶自相似过程，且具有自相似系数(又称为Hurst参数，简记为H)$H=1-\frac{\beta}{2}$，$0<\beta<1$。

自相似系数是描述自相似特征的唯一参数，且有3个表示不同物理意义的取值范围：即当 $0<H<0.5$ 时，表示负相关；$H=0.5$ 时，表示短相关(short-range denpendent，SRD)；$0.5<H<1$ 时，表示长相关(即有自相似性)，且 H 越大，自相似程度越高。这种短相关过程与长相关过程之间的区别就是所谓的Hurst效应。

引理1 对于一个协方差平稳的离散随机过程 $X=(X_t:t=1,2,3,\cdots)$，有

(1)X_t 的自相关函数满足：

$$r(k)=\frac{1}{2}[(k+1)^{2-\beta}-2k^{2-\beta}+(k-1)^{2-\beta}]$$

(2)X_t 的 m 阶叠加过程满足：

$$\mathrm{var}(X^{(m)})=\sigma^2m^{-\beta}$$

(3)X_t 的谱密度函数满足：

$$S(f)=c|e^{2\pi if}-1|^2\sum_{l=-\infty}^{+\infty}\frac{1}{|f+l|^{3-\beta}}$$

若 X_t 满足3个条件中的任意1个，则 X_t 是自相似的，且这3个命题等价[14,20]。

自相似过程的最重要特征是：当 $m\to\infty$ 时，其聚集过程 $X_t^{(m)}$ 的自相关结构是非退化的。传统负载模型则不同，当 $m\to\infty$ 时，聚集过程 $X_t^{(m)}$ 的自相关结构将会退化，即 $r^{(m)}(k)\to 0(k=1,2,3,\cdots)$。

3 Hurst参数估计

对自相似性进行量化的主要手段是Hurst参数(通常表示为H)。由于Hurst参数在自相似I/O负载建模中的作用举足轻重，因此准确估计Hurst参数显得非常重要。Karagiannis等人开发了一种软件SELFIS[21,22]，专门用来估计Hurst参数。Hurst参数的估计方法通常可以分为两类：一类在时域操作，称为时域估值器，包括R/S分析(Rescaled Adjusted Range)[8]和方差时间图(Variance-time plot)[12]等；另一类在频域或小波域操作，称为频域或小波域估值器，包括周期图(Periodogram)[13]、Whittle估值器[20]等。其中，最早也最常用的估计方法是R/S分析。上述均是常用的Hurst参数估计方法，下面分别对其进行介绍。

3.1 R/S分析

R/S分析法又称为Pox图或重标定域分析，被广泛用于数据流量自相似性的参数估计[8,20]，是一种精度比较高的Hurst参数估算方法。下面对其基本思想进行介绍。

对于一组观测值 $\{X(t),1\leqslant t\leqslant n\}$，其样本均值为 $\overline{X}(n)$，样本方差为 $S^2(n)$，相应的R/S统计量可以定义为：

$$\frac{R(n)}{S(n)}=\frac{1}{S(n)}[\max(0,W_1,W_2,\cdots,W_n)-\min(0,W_1,W_2,\cdots,W_n)] \tag{2}$$

式中，$W_k=\sum_{i=1}^{k}X(i)-k\overline{X}(n)$，$1\leqslant k\leqslant n$。

于是可以得到：

$$E\left[\frac{R(n)}{S(n)}\right]\sim cn^H,n\to\infty \tag{3}$$

式中，c 为独立于 n 的正数，H 为Hurst参数，且 $0.5<H<1$。

如果一组实际观测值的长度为 N，将这一组观测值分成长度为 N/K 的 K 个子集，且各个子集互不相交。于是，R/S统计量 $R(t_i,n)/S(t_i,n)$ 可以通过这 K 个互不相交的子集的起始点 t_i 来计算。其中，$t_i=[N/K](i-1)+1$([x]表示不大于 x 的最大整数)标识每个子集的起始时间，且 n 满足关系式 $t_i-1+n\geqslant N$。

对式(3)两边取对数,得:

$$\log(E\left[\frac{R(n)}{S(n)}\right])\sim H\log(n)+\log(c),n\to\infty \quad (4)$$

于是,对于不断变化的 n 值,可以通过描绘出 $\log(E\left[\frac{R(n)}{S(n)}\right])$ 与 $\log n$ 的关系图来估计 Hurst 参数 H,这就是著名的 Pox 图。

R/S 分析法可以准确判断存储负载是否具有自相似性,并能估计相应的 Hurst 参数来反映 I/O 负载的自相似程度,且对 I/O 负载的重尾特征具有健壮性。但是,R/S 分析法只有对足够大的样本空间进行分析时才非常有效。用它来分析小样本数据,得到的分析结果并不可靠,因为 R/S 分析法对短相关性结构非常敏感[20]。

3.2 方差时间图

另一种估计自相似参数的常用方法是方差时间图[12,20]。由文献[12]可知其基本思想是:将 I/O 样本数据序列 $X=\{X(t)|X(1),X(2),\cdots,X(n)\}$ 等分成长度为 n/m 的 m 个子数据集。样本序列 X 的 m 阶聚集过程 $X^{(m)}$ 可以表示为:

$$X^{(m)}(t)=\frac{1}{m}\sum_{i=0}^{m-1}X(tm-i) \quad (5)$$

根据式(5)计算 $X^{(m)}(t),t=1,2,\cdots,n/m$。

由引理 1 可知,样本序列 X 的 m 阶聚集过程 $X^{(m)}$ 的方差满足关系式:

$$\mathrm{var}(X^{(m)})=\sigma^2 m^{-\beta},m\to\infty \quad (6)$$

在式(6)两边取对数,得:

$$\log(\mathrm{var}(X^{(m)}))=\log\sigma^2-\beta\log m,m\to\infty \quad (7)$$

由于 $\log\sigma^2$ 是与 m 无关的常数,如果将 $\mathrm{var}(X^{(m)})$ 作为 m 的函数,可得到一个数对$(\log(m),\log(\mathrm{var}(X^{(m)})))$,并画出 $\log(\mathrm{var}(X^{(m)}))$ 与 $\log m$ 的关系图,得到一条斜率为 $\beta=-2(1-H)$ 的直线,进而根据 β 与 H 的关系式求出 H。

同 R/S 分析法一样,方差时间图被广泛用于数据流量自相似性的参数估计。方差时间图的特点是能比较直观地判断样本负载是否具有自相似性,且计算复杂度较低,但健壮性比 R/S 分析法差。

3.3 周期图

周期图是基于频域的一种估计方法,也是一种图形化估计方法。由文献[13]可知:对于 I/O 样本数据序列 $X=\{X(t)|X(1),X(2),\cdots,X(n)\}$,在离散时间上的谱密度可通过在时间周期 n 上的傅立叶级数获得估计值[13,23]:

$$I(\omega)=\frac{1}{2\pi N}\left|\sum_{j=1}^{n}X(j)e^{ij\omega}\right|^2 \quad (8)$$

式中,ω 表示频率,n 表示时间周期长度或 I/O 样本数据序列中元素的个数。这种估计在文献[13]中被称为周期图。如果样本数据序列 X 为长相关序列,则该序列的谱密度 $I(\omega)$ 与 $|\omega|^{1-2H}$ 成正比。此时,$I(\omega)$ 的对数与频率绝对值 $|\omega|$ 的对数存在线性关系。于是,可以得到一个周期图与频率的 log-log 图,该图是一条斜率为 $\beta=1-2H$ 的直线,从而可以计算出 Hurst 参数的大小。

3.4 Whittle 估计法

Whittle 估计法同样是基于频域的,但它是一种非图形化估计方法。假定观察到的时间序列是参数为 H 的一个自相似随机过程,譬如分数布朗运动过程,且具有特定的形式,那么该过程在频率 ω 处的谱密度就可以表达为 $S(\omega,H)$,其中密度的形式是已知的,但参数 H 是未知的。定义下列函数:

$$Q(H)=\int_{-\pi}^{\pi}\frac{I(\omega)}{S(\omega,H)}d\omega \quad (9)$$

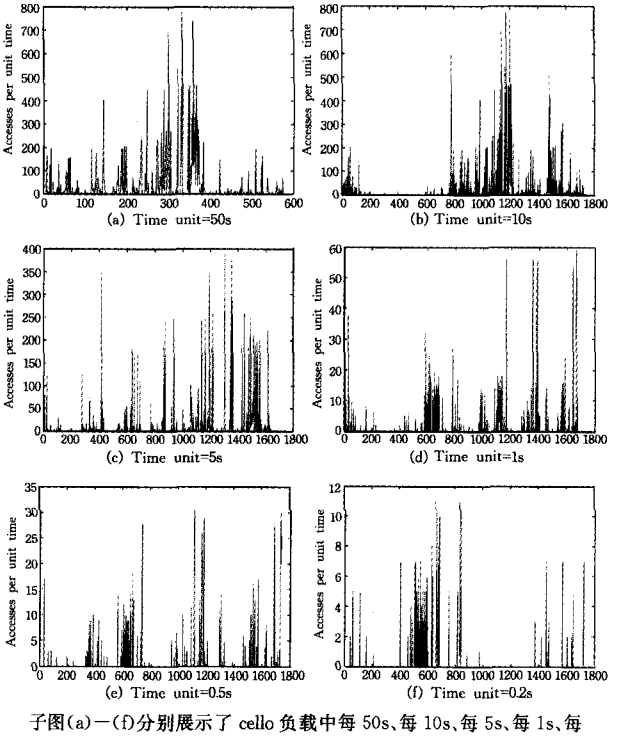
式中,$I(\omega)$ 由式(8)表示,$S(\omega,H)$ 为在频率 ω 处的谱密度。采用极大似然估计,在式(9)取值最小化的情况下,可以确定 H 的大小[23]。

Whittle 估计法不仅可以对 I/O 样本数据进行精确的统计分析,还可以对一段短时样本数据进行分析,求出 H 的估值并给出相应的置信区间。其缺点在于,计算复杂度较高。

纵观上述 4 种估计方法,基于时域的方差时间图及 R/S 方法这两种技术更适合用来检验一个时间序列是否具有自相似性。如果确认该时间序列是自相似的,可以采用 Whittle 估计对 H 值作较为精确的估计。除了上述常用的 Hurst 参数估计算法外,还有一些其它的 Hurst 参数估计算法,如基于时域的绝对值-时间法、Higuchi 方法、回归残数法等[23],基于频域的小波分析法等[23],详细介绍见文献[23]。尽管可以通过不同的方法来估计 Hurst 参数,但其均难以对 Hurst 参数估计值给出权威结论[24,25]。

4 I/O 负载自相似性

对于 I/O 负载来说,自相似性的主要表现是:尽管时间尺度发生变化,但 I/O 负载所具有的突发性、突发聚集性等特征能够得到保持,且这种保持过程也具有非常相似的特点。存储负载中的 I/O 访问序列没有固定的突发长度:突发区间由一些突发子区间构成,这些突发子区间又由一些更小的突发子区间构成。这就是存储系统 I/O 负载的自相似性。下面采用 Gomez[15] 的分析结论详细说明之。



子图(a)—(f)分别展示了 cello 负载中每 50s、每 10s、每 5s、每 1s、每 0.5s 和每 0.2s 的 I/O 到达数[15]。

图 1 不同时间尺度下的 I/O 到达数

Gomez[15] 通过对 cello trace 的分析,研究了磁盘 I/O 访问的自相似性,图 1 直观展现了 cello I/O 负载中的自相似特性,即 cello trace 中单位时间 I/O 到达数的变化曲线。其中,横轴表示 I/O trace 数据的时间,图 1(a)—(f)的时间单位分别是 50s、10s、5s、1s、0.5s 和 0.2s,纵轴表示单位时间内的 I/O 到

达数。对于图1中的某个子图，紧邻其后的子图是从前一个子图中提取的一个子集，即下一幅子图是从上一幅子图中随机选取的一个子区间并将时间分辨率提高之后得到的。

由图1可知，I/O突发没有固定的长度：在每一个时间尺度上，I/O突发都是由突发较大的小区间和突发较小的小区间交替组成的。这说明，I/O突发行为所在的时间范围由包含具有类似突发行为的子区间组成，而这样的子区间又由包含具有类似突发行为的更小子区间组成。也就是说，在不同时间尺度下的I/O到达均体现出明显的突发行为。因此，图1提供了I/O trace自相似性的图示化证据。

然而，传统的泊松模型在较大时间尺度上严重低估了I/O负载的突发程度。由图1所展现出来的在各个时间尺度上都存在的I/O突发性，是自相似现象的一个强有力标志。与基于泊松假定的传统模型相比，自相似负载模型常常更适合描述各种存储负载中的I/O突发行为。因而有必要讨论以仿真生成、测量和数值方法等为主的I/O负载合成方法，用以生成测试基准程序、模拟真实负载来评测存储系统。

5 I/O负载合成模型

目前，已有许多方法被用来生成自相似业务，分别包括：(1)自相似模型，如ON/OFF模型[13,18]、$M/G/\infty$排队模型(又称Cox模型)[26]、分形更新叠加过程(Superposition of Fractal Renewal Process，Sup-FRP)[27]、随机中点置换法(Random Midpoint Displacement，RMD)[28]、分形布朗运动(Fractional Brownian Motion，FBM)[29]、分形高斯噪声(Fractional Gaussian Noise，FGN)[30]、伪自相似法(pseudo-self-similar)[9]和分形自回归滑动平均(Fractional Auto-regressive Integrated Moving Average，FARIMA)[8,30]等；(2)多分形模型，如二项式多分形(Binomial Multifractals)模型[31]，以及基于二项式多分形的b模型[16]和PQRS模型[32]等；(3)基于alpha稳定过程的负载模型[33]。其中，被不少研究者采用的FBM模型虽然易于处理，参数简约，且能在高斯条件下描述自相似性，但是无法同时描述长相关性(LRD)和短相关性(SRD)[29]。而FARIMA模型虽极为灵活，能够同时描述长相关性和短相关性，但过于复杂，仿真运算量太大，且对负载中的突发性缺乏表述。例如，Gartett和Willinger[30]采用FARIMA模型合成的VBR视频流，难以反映出实际VBR流中的突发性。

虽然以上负载模型各有其优缺点，但这些方法同样适用于对I/O负载的描述。目前，已有部分方法如ON/OFF模型、Cox模型、Sup-FRP模型、基于二项式多分形的负载模型和基于alpha稳定过程的负载模型被运用到I/O负载的描述中，下面着重对其进行介绍。

5.1 ON/OFF模型

ON/OFF模型能对自相似现象给出明确的物理解释，有助于理解产生自相似现象的原因，被Gomez[14,15,34]、Gribble[17]分别应用于块级I/O负载和文件级I/O负载的合成研究中。在理论上，ON/OFF模型是将整个I/O负载视为许多独立同分布的ON/OFF过程的叠加。一个ON/OFF过程可以表示为在一对信源和目的地之间传输的数据业务流。整个ON/OFF过程是由两个严格交替的周期组成：ON周期和OFF周期。ON周期和OFF周期呈现出诺亚效应(Noah Effect)，即有高可变性或无限方差。将许多ON/OFF源叠加，就能生成呈现出约瑟夫效应(Joseph Effect，即长相关性)的聚集负载。流叠加法将自相似过程看成是无数用户数据源叠加的结果，可利用ON/OFF业务源模拟用户数据源产生自相似过程。产生原理如下：ON/OFF模型只有ON和OFF两个状态，当业务处于ON状态时以恒定的速率产生数据，而处于OFF状态时不发送任何数据。如果源的负载出现了包含高可变长度的ON期和OFF期，单个源就被认为是一个ON/OFF源。ON期存在很多活动，而OFF期则完全缺少活动。

诺亚效应反映了ON期和OFF期的高可变性。因此，如果各个数据源的ON周期和OFF周期长度分布是重尾的，那么将许多数据源产生的I/O负载过程叠加在一起而生成的I/O负载将会是(渐进)自相似的，这与实际测量结果是一致的。如果ON期和OFF期长度分布都是重尾的，则对于区间长度T，它们满足$P(T>t)\sim at^{-\alpha}$，$t\to\infty$，$1<\alpha<2$。如果数据源在ON期的活动是一致的，则将多个此类数据源叠加就得到自相似过程(自相似参数为$H=(3-\alpha)/2$)。如果分别采用参数为α_1和α_2的Pareto分布作为ON阶段和OFF阶段的过程分布，叠加的结果是以$H=(3-\min(\alpha_1,\alpha_2))/2$为自相似参数的分形高斯噪声过程。

通过将许多呈现出诺亚效应的ON/OFF源进行叠加，便可解释I/O到达模式的自相似特性[18]。究其原因，I/O负载中存在的自相似性与存储在磁盘中的文件大小有关，而在众多的存储系统中，文件大小分布是重尾的[13]，那么访问文件所需要的时间也应服从重尾分布。因此，对应于文件的读或写，相应的I/O访问行为可以通过重尾分布进行刻画。在合成负载过程中，通过选取源的数目及Pareto分布的形状参数α，便可调整访问数量[35]。

ON/OFF模型构造简单，且构造过程具有明确的物理意义。这使得ON/OFF模型能从构造过程的因果机理上解释自相似现象[36]。但是，在构造ON/OFF模型的过程中作了很多前提假设，且这些前提假设条件常常与实际情况不相符合。这使得ON/OFF模型难以对实际负载进行预测。

5.2 $M/G/\infty$排队模型

Gomez等人[14]研究发现，ON/OFF模型虽适合描述持续时间较长的源的自相似现象，但难以有效描述持续时间较短的源的自相似现象。于是，文献[14]提出$M/G/\infty$排队模型来弥补这一不足。

在$M/G/\infty$排队模型中，I/O到达过程被假设为均值为ρ的泊松过程。令$\{X(t),t=0,1,2,\cdots\}$为$M/G/\infty$排队模型中I/O到达过程在时刻t的计数过程，I/O响应时间分布函数为$F(x)$，则$X(t)$的自相关函数为：

$$r(k)=\rho\int_k^{\infty}(1-F(x))\mathrm{d}x \tag{10}$$

如果I/O响应时间分布函数为Pareto分布，则：

$$r(k)=\rho\int_k^{\infty}(\frac{\alpha}{k})^{\beta}\mathrm{d}x=\frac{\rho\,\alpha^{\beta}}{\beta-1}k^{1-\beta} \tag{11}$$

那么，计数过程$X(t)$是渐进自相似过程，且自相似参数为$H=(3-\alpha)/2$。虽然研究表明，$M/G/\infty$排队模型能准确描述持续时间较短的源的自相似现象，但该方法需要在计算量和计算精度之间进行折中。

5.3 分形更新叠加过程

Hsu[6]指出，将独立同分布的分形更新过程(Fractal Renewal Process，FRP)进行叠加是生成I/O访问序列的一种有效途径，具体算法实现见文献[6]，本文仅对其进行简要介绍。

在文献[6]中，Sup-FRP模型被定义为M个独立且概率同分布的FRP的叠加。由于每个I/O到达过程(即FRP流)是一个更新点过程，Sup-FRP模型完全由分形更新过程的个数M和一般到达间隔概率密度函数(probability density function, pdf)($pdf>p(\tau)$)来决定。用$\tau_i^{(j)}$表示以概率密度$p(\tau)$到达的第j个I/O到达过程的第i个到达间隔时间。概率密度函数为[37]：

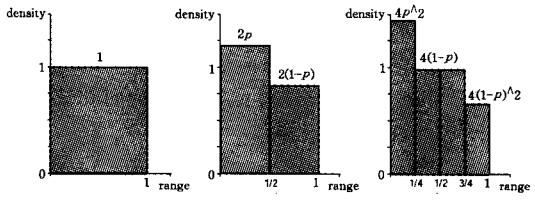
$$p(\tau)=\begin{cases}A^{-1}\gamma e^{-\pi A^{-1}}, & \tau\leqslant A\\ \gamma e^{-\gamma}A\gamma\tau^{-(\gamma+1)}, & \tau>A\end{cases} \tag{12}$$

式中，$\gamma=2-\alpha$。当$\tau\leqslant A$时，$p(\tau)$以指数函数衰减；当$\tau>A$时，$p(\tau)$以幂函数衰减。事实上，由式(12)可知，由于$p(\tau)$的“重尾”，具有分形行为的Sup-FRP模型的业务源的到达间隔时间非常大(以很大的概率跨越多个数量级)。这样一来，Sup-FRP模型在$A\ll T$范围表现出分形行为[37]。Sup-FRP模型具有3个参数：$\alpha(0<\alpha<1)$、A和M，其中参数M控制着Sup-FRP模型的突发性，而Hurst参数为$H=\frac{(\alpha+1)}{2}$。

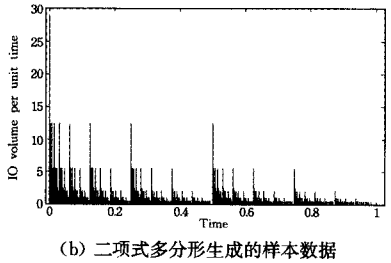
5.4 二项式多分形

对于自相似I/O负载，Hurst参数能从长时间范围描述其突发性，但磁盘I/O在短时间范围内的局部突发对存储系统性能的影响更大。由于适用于描述大时间尺度上I/O突发行为的传统自相似模型，通常难以准确描述小时间尺度上的I/O突发行为，Wang[16,32]、Hong[31]等人用二项式多分形来描述I/O到达过程。下面对二项式多分形的构造机理进行简要介绍。

多分形其实就是单分形自相似过程在时间相关尺度下的扩展。对于测度为1的单位区间$I=[0,1]$，首先将I分成两个等长的子区间$I_0=[0,1/2]$(测度为$m_0=p>1/2$)和$I_1=[1/2,1]$(测度为$m_1=1-m_0=1-p$)，I_0和I_1的概率密度分别是$2p$和$2(1-p)$。然后，将I_0分成两个等长的子区间$I_{00}=[0,1/4]$(测度为$m_{00}=p^2$)和$I_{01}=[1/4,1/2]$(测度为$m_{01}=p(1-p)$)，I_{00}和I_{01}的概率密度分别是$4p^2$和$4p(1-p)$；将I_1分成两个等长的子区间$I_{10}=[1/2,3/4]$(测度为$m_{10}=p(1-p)$)和$I_{11}=[3/4,1]$(测度为$m_{11}=(1-p)^2$)，I_{10}和I_{11}的概率密度分别是$4p(1-p)$和$4(1-p)^2$。依此递归，即可构造出一个二项式多分形[16,31]，如图2所示。



(a) 二项式多分形生成过程



(b) 二项式多分形生成的样本数据

图2 二项式多分形的生成过程[31]

对$\forall x\in[0,1)$，存在唯一的子区间$I_{\varepsilon_1\varepsilon_2\cdots\varepsilon_n}$包含$x$，记为$I^{(n)}(x)$。为方便起见，假定$m_0>m_1$，对于$x$，$I^{(n)}(x)$的密度定义为：

$$\frac{\mu(I^{(n)}(x))}{|I^{(n)}(x)|}=\frac{m_{\varepsilon_1}m_{\varepsilon_2}\cdots m_{\varepsilon_n}}{2^{-n}} \tag{13}$$

当$n\to\infty$时，$I^{(n)}(x)$趋于无穷，$\mu(I^{(n)}(x))$是$I^{(n)}(x)$的测度，$|I^{(n)}(x)|$表示$I^{(n)}(x)$的长度。以0.5的概率随机选取m_0或m_1作为左乘子，得到的二项式多分形与实际I/O负载中的突发情况更接近。

局部时间尺度指数$\alpha(x)$(也称Holder指数)和二项式测度的多分形谱$f(\alpha)$被用来描述小时间尺度上的I/O突发。Holder指数可表示为：

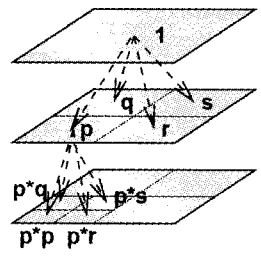
$$\alpha(x)=\lim_{n\to\infty}\frac{\log_2\mu(I^{(n)}(x))}{\log_2|I^{(n)}(x)|} \tag{14}$$

式中，对于时刻x_0，$\alpha(x_0)<1$表示相应时刻的局部突发性较大；$\alpha(x_0)>1$表示相应时刻的局部变化程度较低；若对任意时刻x_0，$\alpha(x_0)=H$，相应过程为单分形，即参数为H的严格自相似过程，否则为多分形。二项式测度的多分形谱可表示为：

$$f(\alpha)=\lim_{n\to\infty}f^{(n)}(\alpha)=\lim_{n\to\infty}\frac{\log_2 N^{(n)}(\alpha)}{n} \tag{15}$$

式中，$N^{(n)}(\alpha)$为Holder指数值为α的子区间$I^{(n)}$的个数。$f(\alpha)$具有与熵函数一样的形式，提供了估计m_0的一条途径(偏差为p)。二项式多分形里的偏差参数p反映了局部突发行为，这种偏差参数能从实际I/O trace中评估出来。Wang[16]等人以此构造了描述I/O到达模式的b模型，具体算法实现见文献[16]。

基于二维的二项式多分形，Wang[32]等人对b模型进行二维扩展，得到PQRS模型，它从时空上对I/O到达模式和访问模式进行描述，如图3所示。在图3中，横轴表示I/O到达时刻，纵轴表示I/O访问的逻辑块地址(Logical Block Address, LBA)。显然，一个I/O请求落入磁盘时空图(顶层方格)的概率是1。采用类似于二项式多分形构造的方法，首先从二维上将时空图均分成4个方格，I/O请求落入各个方格的概率分别是p、q、r和s，且有$p+q+r+s=1$。然后，将这4个方格分别均分成4个更小的方格，I/O请求落入各个方格的概率如图3(a)所示。依此递归，即可构造出二维的二项式多分形以描述I/O负载的二维时空图，具体算法实现见文献[32]。



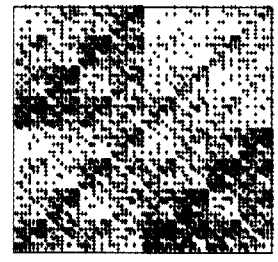


图3 PQRS模型的生成过程[32]

由以上内容可知，二项式多分形以及基于二项式多分形的b模型和PQRS模型具有参数少的优点，易于构造。特别是b模型，仅仅依赖一个参数，即偏差p，便可从实际trace中估计得到。此外，PQRS模型仅适用于高斯条件下的自相似I/O负载合成。

5.5 Alpha稳定过程

对多组典型磁盘负载的研究结果表明[33]：一方面，由于应用类型不同，有的存储负载中的I/O到达是独立同分布的，

有的存储负载则具有自相似性；另一方面，即使同样具有自相似性，有的I/O负载具有高斯性，而有的I/O负载则体现出非高斯性。Zou等人[33]提出了一种基于alpha稳定过程的I/O负载模型，它既能有效刻画自相似I/O负载中的高斯性和非高斯性，又能精确模拟各种应用负载中的I/O突发行为。下面对alpha稳定过程理论进行简要介绍。

定义1[33,38] 一个随机变量X被称为具有稳定分布，若存在参数$0<\alpha\leqslant 2,\sigma>0,-1\leqslant\beta\leqslant 1,\mu\in R$，使得其特征函数具有如下形式：

$$E[e^{i\theta X}]=\begin{cases}e^{-\sigma^{\alpha}|\theta|^{\alpha}(1-i\beta sign\theta\tan\frac{\pi\alpha}{2})+i\mu\theta}, & \alpha\neq 1\\ e^{-\sigma|\theta|(1+i\beta sign\theta\ln|\theta|)+i\mu\theta}, & \alpha=1\end{cases} \tag{16}$$

式中，$sign(\cdot)$为示性函数，且有：

$$sign\theta=sign(\theta)=\begin{cases}1, & \theta>0\\ 0, & \theta=0\\ -1, & \theta<0\end{cases} \tag{17}$$

此外，特征指数α表示分布中的突发程度，偏斜参数(the skewness parameter)β表示分布的尾部变化情况。如果$\beta\neq 0$，说明相应的密度函数是偏斜的：取负值表示密度函数偏斜向左尾部(left-tail)；反之，则表示密度函数偏斜向右尾部(right-tail)。由此可见，整个分布函数的基本形状是由参数α和β决定的。同时，尺度参数(scale parameters)σ表示分布的方差，位置参数(location parameters)μ表示分布的均值。习惯上，将具有上述参数且服从alpha稳定分布的随机变量简记为$X\sim S^{(\alpha)}_{\sigma,\beta,\mu}$[33,38]。

由式(16)可知，当$\alpha=2$时，有：

$$E[e^{i\theta X}]=e^{-\sigma^2\theta^2+i\mu\theta} \tag{18}$$

此时，alpha稳定分布的特征函数退化为高斯特征函数，相应的均值为μ，方差为$2\sigma^2$，β无意义，因为在式(16)中，$\beta\tan\pi=0$恒成立。当$0<\alpha<2$时，式(16)表示的是一种非高斯特征函数。因此，随着α取值的不同，alpha稳定过程可以表示高斯和非高斯情况下的随机过程。

基于alpha稳定过程理论，Zou等人[33]构造了alpha稳定I/O负载模型，相应的形式化表述为：

$$IOs(i)=\nu N_{\alpha,\beta,H}(i)+\delta \tag{19}$$

式中，$IOs(i)$表示第i个单位时间内的I/O请求数，参数ν和δ均是大于0的实数，ν表示I/O负载的平均速率，δ表示I/O负载相对于平均速率的偏差程度(the deviation degree)，α表示I/O负载的突发程度，β表示I/O负载的重尾化程度，H表示I/O负载的自相似程度。此外，$N_{\alpha,\beta,H}(i)$满足以下关系式：

$$N_{\alpha,\beta,H}(i)=\sum_{k=1}^{Km}h_d\left(\frac{k}{m}\right)\cdot S^{\alpha}_{(1+\beta/2)^{1/\alpha},1,0}\left(i-\frac{k}{m}\right)-\sum_{k=1}^{Km}h_d\left(\frac{k}{m}\right)\cdot\tilde{S}^{\alpha}_{(1-\beta/2)^{1/\alpha},1,0}\left(i-\frac{k}{m}\right)$$

其中，$h_d(x)=\begin{cases}x^d-(x-1)^d, & x\geqslant 1\\ x^d, & 0<x\leqslant 1\end{cases}$，$d=H-\frac{1}{\alpha}$，$S^{(\alpha)}_{1,1,0}$和$\tilde{S}^{(\alpha)}_{1,1,0}$表示两个独立同分布的离散随机变量。

在选用alpha稳定过程作为I/O负载建模的依据之前，需要对实际的I/O trace数据进行测量分析，以验证采用alpha稳定过程的合理性。此外，alpha稳定I/O负载模型包含的5个参数，每个均具有相应的物理意义，这使得研究人员可以针对不同的应用环境很方便地转换I/O负载模型，比如，可以通过改变参数α的值来灵活地刻画高斯和非高斯条件下I/O负载中的突发。

6 存在的问题及发展趋势

从存储负载自相似性研究的现状来看，有以下问题尚未完全得到解决：

(1)研究I/O负载的自相似性，其意义在于设计和优化存储子系统，并对系统进行公正和无偏的性能评价。可是，I/O负载的自相似模型在存储子系统的负载预测、I/O性能分析等方面的应用至今仍然不多。

(2)评估存储负载的自相似程度，当前仅能对其作出定性的估计，尚无一种Hurst参数估值器能从定量上对其进行精确估计，给出权威的估计值。

(3)如何根据I/O负载的自身特点建立恰当的负载模型，仍需深入研究。采用负载模型合成自相似I/O负载，其关键是精确描述负载中的I/O突发。已有的自相似I/O负载模型各有优缺点，且通常都缺乏明确的解析表达式，不易应用于存储系统的性能评价和优化。

当前，为了适应海量存储应用的需要，大规模网络存储系统的应用越来越广泛。同时，存储系统的种类多样(如智能存储、融合存储、云存储等)，结构各异，应用模式也变得越来越复杂。在新形势下，前人的I/O负载自相似性研究结果常常难以真实体现如今存储负载中实际的I/O访问情况。因此，当前仍有必要广泛收集和深入研究典型的I/O traces，及时把握当前各种I/O负载的特点，做到与时俱进，以指导海量存储系统的设计和系统性能的改善。

结束语 I/O负载特征研究对改善存储系统性能至关重要，而有效的流量模型则是理解和预测I/O负载行为、分析I/O性能、设计存储系统并对其服务质量进行评价的理论基础。研究表明，实际I/O负载具有自相似性，传统的泊松假定与实际情况不相匹配。本文详细介绍了几种常用的I/O负载自相似参数的估计方法，并对已有的I/O负载合成模型进行了剖析，讨论了各种自相似模型、多分形模型以及alpha稳定模型的特点，探讨了有待解决的问题，并对I/O负载自相似性研究的发展趋势进行了展望，为存储负载的自相似性研究提供了有益参考。

参考文献

[1] Sacks D. Demystifying DAS, SAN, NAS, NAS gateways, Fibre Channel, and iSCSI[R]. Mar. 2001

[2] 覃灵军. 基于对象的主动存储关键技术研究[D]. 武汉：华中科技大学，2007

[3] Gray J. Put EVERYTHING in the storage device[R]. Talk at NASD Workshop on Storage Embedded Computing. June 1998

[4] Hsu W W, Smith A J, Young H C. Projecting the performance of decision support workloads on systems with smart storage (SmartSTOR)[C]// Proceedings of IEEE Seventh International Conference on Parallel and Distributed Systems (ICPADS). Iwate, Japan, July 2000: 417-425

[5] Les F. Outsourced network storage[R]. PC Magazine (What's In Storage for You?), Mar. 2001

[6] Hsu W W. Dynamic Locality Improvement Techniques for Increasing Effective Storage Performance[R]. Technology Report. University of California at Berkeley, 2002

[7] Ganger G. Generating representative synthetic workloads[C]// Proceedings of the Computer Measurement Group Conference. Dec. 1995: 1263-1269

[8] Leland W, Taqqu M, Willinger W, et al. On the self-similar nature of Ethernet traffic (extended version)[J]. IEEE/ACM Transactions on Networking, 1994, 2(1): 1-15

[9] Paxson V, Floyd S. Wide-area traffic: The failure of poisson modeling[J]. IEEE/ACM Transactions on Networking, 1995, 3(3): 226-244

[10] Quan Z, Chung J. Priority queueing analysis of self-similar traffic in high-speed networks[C]// Proceedings of the IEEE International Conference on Communications (ICC). 2003: 1606-1610

[11] Jain R, Routhier S. Packet trains-measurements and a new model for computer network traffic[J]. IEEE J. Select. Areas Common, 1986, 4(6): 986-995

[12] Beran J, Sherman R, Taqqu M S, et al. Long- range dependence in variable-bit-rate video traffic[J]. IEEE Transactions on Communications, 1995, 43: 1566-1579

[13] Crovella M E, Bestavros A. Self-similarity in World Wide Web traffic: evidence and possible causes[J]. IEEE/ACM Transactions on Networking, 1997, 5: 835-846

[14] Gomez M E, Santonja V. Analysis of Self-Similarity in I/O Workload Using Structural Modeling[C]// Proceedings of the 8th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS). College Park, Maryland, 1999: 234-242

[15] Gomez M E, Santonja V. Self-similarity in I/O workload: Analysis and modeling[C]// Proceedings of the Workshop on Workload Characte- rization (held in conjunction with the 31st annual ACM/IEEE International Symposiumon Microarchitecture). Dallas, Texas, Nov. 1998: 97-104

[16] Wang M, Madhyastha T. Data mining meets performance evaluation fast algorithm for modeling bursty traffic[C]// Proceedings of the 18th International Conference on Data Engineering (ICDE). Rome, Italy, Feb. 2002: 507-516

[17] Gribble S, Manku G, Brewer E. Self-similarity in high-level file systems: Measurement and applications[C]// Proceedings of the ACM SIGMETRICS. Madison, Wisconsin, June 1998: 141-150

[18] Gomez M E, Santonja V. A new Approach in the Modeling and Generation of Synthetic Disk Workload[C]// Proceedings of the 9th International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems (MASCOTS). 2000: 199-206

[19] 李明强，舒继武. I/O负载自相似研究综述[J]. 计算机研究与发展，2008，45(6)：1072-1084

[20] Willinger W, Taqqu M, Sherman R. Self-similar Through High-variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level[J]. IEEE/ACM Transactions on Networkingy, 1997, 5(1): 71-86

[21] Karagiannis T, Faloutsos M. SELFIS: A Tool for Self-similarity and Long-Range Dependence Analysis[C]// Proceedings of the 1st Workshop on Fractals and Self-similarity in Data Mining: Issues and Approaches. Edmonton, Canada, 2002

[22] Karagiannis T, Faloutsos M, Molle M. A User-friendly Self-similarity Analysis Tool[J]. ACM SIGCOMM Computer Communication Review, 2003, 33(3): 81-93

[23] 洪飞. 基于小波的网络业务研究[D]. 北京：中国科学院研究生院，2004

[24] Karagiannis T, Molle M, Faloutsos M. Long-range Dependence: Ten Years of Internet Traffic Modeling[J]. IEEE Computer Society, 2004, 8(5): 57-64

[25] Karagiannis T, Faloutsos M, Riedi R. Longrange Dependence: Now You See It, Now You Don't[C]// Proceedings of Globecom. Taiwan, 2002

[26] Breslau L, Cao P, Fan L, et al. Web Caching and Zipflike Distributions: Evidence and Implications[C]// Proceedings of the IEEE Infocom Conference. March 1999: 126-134

[27] Ruy B. Real-time Generation of Fractal ATM Traffic: Model, Algorithm and Implementation[R]. Center for Telecommunications Research, Columbia University, 1996

[28] Lau W, Erramilli A, Wang J, et al. Selfsimilar Traffic Generation: The Random Midpoint Displacement Algorithm and Its Properties[C]// Proceedings of the IEEE International Conference on Communications (ICC). Seattle, WA, 1995

[29] Norros. On the use of fractional Brownian motion in the theory of connectionless networks[J]. IEEE J. Selected Areas Commun, 1995, 13(6): 953-962

[30] Garrett M, Willinger W. Analysis, modeling and generation of self-similar VBR video traffic[C]// Proceedings of the SIGCOMM. 1994: 269-280

[31] Hong B, Madhyastha T. The relevance of long-range dependence in disk traffic and implications for trace synthesis[C]// Proceedings of the 22nd IEEE/ 13th NASA Goddard Conference on Mass Storage Systems and Technologies. 2005: 316-326

[32] Wang M, Ailamaki A, Faloutsos C. Capturing the spatio-temporal behavior of real traffic data[C]// Proceedings of the IFIP International Symposium on Computer Performance Modeling, Measurement, and Evaluation. Rome, Italy, 2002: 23-27

[33] Zou Q, Feng D, Zhu Y, et al. A Novel and Generic Model for Synthesizing Disk I/O Traffic Based on The Alpha-stable Process[C]// Proceedings of the 16th Annual Meeting of the IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems. Baltimore, MD, USA, September 2008: 1-10

[34] Gomez M E, Santonja V. An new approach in the analysis and modeling of disk access patterns[C]// Proceedings of IEEE International Symposium on Performance Analysis of Systems and Software. Austin, Texas, 2000: 172-177

[35] 黄丽亚，王锁萍. 基于自相似业务流的Hurst加权随机早检测算法[J]. 通信学报，2007，28(4)：95-100

[36] Embrechts P, Maejima M. Self-similar Processes[M]. Princeton University Press, 2002

[37] Ryu B K, Lowen S B. Point process approaches to the modeling and analysis of self-similar traffic: Part I-Model construction[C]// Proceedings of the IEEE INFOCOM. San Francisco, 1996: 65-72

[38] Samorodnitsky G, Taqqu M. Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance[M]. Chapmanand Hall, New York, 1994