磁盘存储系统节能技术研究综述

田 磊 冯 丹 岳银亮 吴素贞 毛 波
(华中科技大学计算机科学与技术学院武汉光电国家实验室 武汉 430074)

摘 要 目前磁盘是构成存储系统的重要组成部分,在存储系统总能耗中磁盘能耗占了大部分。因此磁盘存储系统的高能耗问题受到越来越多研究人员的关注。综述了磁盘存储系统从磁盘到存储系统各个层次的能耗研究进展和现状,同时对各种典型节能方法从原理、实现机制和评测手段等诸方面进行了分析和讨论,并对比分析和总结了各种节能技术的适应环境。结合海量存储系统负载特征的复杂性和应用环境的复杂性等特点,指出了磁盘存储系统节能技术的未来研究方向。

关键词 磁盘存储系统,磁盘阵列,节能技术,能耗测量
中图法分类号 TP311 **文献标识码** A

Survey on Power-saving Technologies for Disk-based Storage Systems

TIAN Lei FENG Dan YUE Yin-liang WU Su-zhen MAO Bo
(Wuhan National Lab for Optoelectronics, College of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract Hard disk drives have been one of important components of modern-day storage systems, and their power consumption is the major part of total power consumption of storage systems. Therefore, the problems of high power dissipation of disk-based storage systems are paid much more attentions by researchers. The research progress and status on power consumption issues from disk to storage system were extensively studied in this paper. Then representative power-saving technologies, implementation mechanisms as well as evaluation methodologies were presented and discussed in details. Further, their respective characteristics and applicability were also analyzed and summarized. By taking workload characteristics of mass storage systems and complexity of application scenarios into consideration, the future work on power-saving technologies for disk-based storage systems was pointed out finally.

Keywords Disk-based storage systems, RAID, Power-saving technologies, Power measurement

1 引言

信息技术的迅猛发展使得数字信息呈爆炸式增长,新产生的数字信息以每年30%的速度递增,磁记录方式以其特有的优势成为当今信息社会不可或缺的数字信息记录方式,高可靠、大容量、高性能的磁盘存储系统成为当前学术界和工业界研究的热点。大规模磁盘存储系统将大量磁盘组合起来并使用各种方法来提高存储系统的可靠性、性能和容量,但磁盘的高能耗已变成影响存储系统运行成本和数据可靠性的关键因素之一。如何有效降低磁盘存储系统的能耗,成为摆在存储研究人员面前的一个重要问题。

和电池驱动的便携式移动存储设备一样,数据中心也非常关注磁盘存储系统的能耗问题。对于一个有多个Web服务器的数据中心,能量消耗量可达每平方英尺75瓦到200瓦。而Google公司的首席执行官Eric Schmidt也表示[4]:"最影响Google设计人员的不是(计算机系统的)速度而是能耗,因为一个数据中心能消耗一座城市的电力。"除此以外,关注磁盘能耗的主要原因还有以下两点:

(1)冷却上的考虑。存储设备的高能耗同时也带来另一个负面效应——高热量。如果磁盘运行时的温度比环境正常温度高5℃,磁盘失效的可能性就会增加10%~15%。对于数据中心这样存储设备密集的环境,为了保障存储系统的可靠性,必须使用额外的降温装置来冷却存储系统,这无疑又会增加数据中心的总能耗和总成本。

(2)总体拥有成本上的考虑。2005年,美国国家数据中心的能耗花费达到40亿美元,而且每年预计增长25%。考虑到磁盘存储部分的能耗占数据中心的能量总消耗量的27%,所以如何节约磁盘的能耗成为数据中心的首要问题。

近年来,磁盘存储系统的能耗问题引起了国外大学和研究机构的广泛关注和研究兴趣,并已取得较丰硕的成果。但

到稿日期:2009-10-26 返修日期:2010-01-09 本文受国家863计划项目(2009AA01A401,2009AA01A402),教育部创新团队(IRT0725),国家自然科学基金(60873028)和国家重点自然科学基金(60933002)资助。

田 磊(1978-),男,博士,讲师,主要研究方向为磁盘阵列等,E-mail:ltian@hust.edu.cn;冯 丹 女,教授,博士生导师,CCF会员,主要研究方向为对象存储系统等;岳银亮 男,博士生,主要研究方向为高能效存储系统等;吴素贞 女,博士生,主要研究方向为存储系统可靠性等;毛 波 男,博士生,主要研究方向为混合存储系统等。

在国内，对存储系统能耗问题的研究还比较少。深入研究存储系统的能耗问题和节能技术有利于设计与实现有效的磁盘存储系统，并有效地提高系统的可靠性和降低运行成本。本文从能耗问题和节能技术演化的角度，针对当前磁盘存储系统能耗研究缺少系统总结的情况，较全面地概括、分析和研究了能耗问题和节能技术的起源、现状和发展方向。

本文第1节介绍存储系统中磁盘内部工作原理和能耗模式；第2节给出磁盘能耗的评价标准及测量方法；第3节介绍磁盘存储系统能耗模型及仿真；第4节介绍磁盘系统节能途径；第5节介绍磁盘系统节能技术的主要进展；最后为全文总结及展望。

2 磁盘工作原理及能耗

一个典型的磁盘由如下部件构成：磁盘片、磁头、磁头臂、永磁铁、音圈马达、主轴和空气过滤片等。磁盘响应一个I/O请求，一般可分为以下4个阶段：(1)寻道阶段：磁头移动到对应的柱面上；(2)旋转阶段：等待盘片旋转到对应的位置上；(3)数据传输阶段：磁头从盘片上读取数据或写入数据到盘片上；(4)空闲阶段：当前I/O请求响应完毕后到开始响应下一个I/O请求之间的阶段。

如图1所示，磁盘的工作状态大致可分为下列3个主要状态：活动状态、空闲状态和待机状态。当磁盘处于活动状态时，盘片高速旋转，磁头同时也在寻道、定位或存取数据，此时磁盘的能耗最大；当磁盘处于空闲状态时，盘片保持旋转状态，磁头臂停止运转，其他大多数电子器件处于关闭状态，此时磁盘的能耗较其处于活动状态时稍低；而当磁盘处于待机状态时，除电子器件关闭外，盘片也停止旋转，磁头归位，此时磁盘的能耗最低。但磁盘从待机状态返回到数据存取状态所需的时间长达数秒(实际时间的长度依磁盘的不同而有所差异)。在磁盘的各个耗能部件中，诸如盘片、磁头臂等机械部分耗能最多。表1给出3种不同应用环境的典型磁盘的技术和能耗参数。

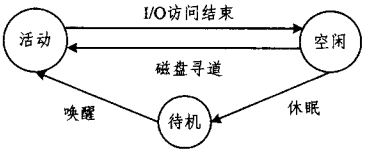

图1 磁盘的3个工作状态

表1 3种典型磁盘的技术和能耗参数[16]

磁盘模型		企业级硬盘	笔记本硬盘	微硬盘
		IBM Ultrastar 36Z15	Toshiba MK5002MPL	IBM DSCM-11000
磁盘物理参数	容量	18.4GB	5GB	1GB
	转速	15000RPM	4000RPM	3600RPM
	平均寻道延迟	3.4ms	15ms	1ms
	平均旋转延迟	2ms	26ms	20ms
	持续传输率	55MB/s	66.7MB/s	4.3MB/s
	功率(活动/空闲/待机)	13.5W/10.2W/2.5W	1.2W/0.9W/0.2W	0.6W/0.5W/0.06W
磁盘能耗参数	休眠/唤醒能量	13J/135J	N/A	N/A
	休眠/唤醒耗时	1.5s/10.9s	15s	2s

3 磁盘能耗的评价标准及测量方法

3.1 磁盘能耗评价标准

磁盘存储系统能量消耗的实际数值是评价磁盘存储系统能耗的衡量标准。能耗指标可分为下列两种：总能耗和每次I/O的平均能耗。总能耗指一段时间内磁盘或磁盘存储系统消耗的能量总和，衡量标准为能量的单位焦耳；而每次I/O的平均能耗是指单次I/O消耗的能量，即总能耗和该段时间内处理的I/O总数的商，衡量标准为焦耳/请求。

3.2 磁盘能耗测量方法

如何准确测量存储系统的能耗直接影响到对磁盘存储系统能耗评价的准确程度。磁盘存储系统的能耗测试主要有下面两种方法：实际测量和模拟仿真。使用电流计(或电压计)测试运行时的电流(或电压)，可以取得较为准确的测量结果，但对于大规模的磁盘存储系统，物理测量难以实施；另一方面，由于很多节能技术是基于假想的磁盘内部工作模式的改造(如盘内并行(Intra-Disk Parallelism[3]))，而这些磁盘目前并没有产品化，因此无法使用物理测量的方法。所以，通过磁盘仿真的能耗测量是普遍采用的测试方法，绝大部分能耗研究都采用了模拟仿真的方法。

3.3 磁盘系统能耗模型及仿真

普林斯顿大学J. Zedlewski等人于2003年提出了一个磁盘能耗模型和性能仿真环境Dempsey[16]。它在不需要磁盘制造商提供详细的磁盘技术参数的情况下，可通过精心设计的实验测试方法来自动获取磁盘性能和能耗参数，因此这个能耗仿真器具有很强的应用性。具体说来，Dempsey是在磁盘仿真器DiskSim的基础上开发而成的，根据磁盘能耗模型和特性增加了评估系统能耗的功能模块，将磁盘性能模型和磁盘能耗模型二者有效结合起来。在DiskSim准确评估磁盘性能的前提下，Dempsey能够同时获取磁盘性能和能耗两方面指标。Dempsey还能准确地评估磁盘各种操作的能耗值，比如寻道、旋转、读、写和空闲周期等。

从实验结果上看，Dempsey仿真的能耗值与真实测量的能耗值误差与具体的磁盘有关，在对1GB IBM Microdrive的测试中，平均误差仅为1.8%，最大误差不超过7.5%；而在对5GB Toshiba Type II PC Card HDD的测试中，平均误差不超过3.6%，最大误差不超过6.9%。

由于磁盘阵列已成为当前磁盘存储系统的基本组成部分，因此目前对磁盘存储系统的能耗研究工作大多集中在磁盘阵列上，而众多磁盘存储系统的仿真器是基于类似于Dempsey这样的仿真模型扩展而成的。

4 磁盘系统节能途径

磁盘存储系统的节能途径有两个：(1)降低单个磁盘所消耗的能量；(2)减少消耗能量的磁盘个数。

4.1 降低单个磁盘所消耗的能量

根据磁盘的运行原理及其物理特征，降低单个磁盘所消耗的能量主要从如下3方面着手：细分磁盘活动状态，减少磁

头定位开销和延长磁盘处于空闲状态的周期。

细分磁盘活动状态：

Carrera 等人和 Gurumurthi 等人[2]针对服务器负载下空闲周期较短而使磁盘转入待机状态的问题，提出动态转速磁盘(DRPM)的概念，即将磁盘的盘片旋转速度分为多个速度等级，在系统负载较轻时使磁盘运转在低速旋转状态；而当系统负载变重时，将磁盘相应调整为高速旋转状态。实验表明，动态转速磁盘模型可有效地降低磁盘所消耗的能量，且系统轻负载时的节能效果优于系统重负载时的节能效果。

减少磁头定位开销：

从磁盘的工作原理可知，磁盘的性能和能耗均受制于磁头定位延迟，让磁盘尽可能进行顺序访问是减少磁头定位开销的最有效的方法。Huang 等人[1]为数据创建多个副本并将其存储在文件系统的空闲块上，通过 I/O 调度的方法使用户请求尽可能地顺序访问磁盘上的数据，从而既提高了用户性能又有效地降低了能耗。实验表明，该方法使得每个请求的平均能耗降低了 40%～71%。

延长磁盘处于空闲状态的周期：

延长磁盘处于空闲状态的周期是最常用、最可行的一种磁盘节能技术。在单磁盘上的延长磁盘空闲时间的节能方法也可应用到磁盘存储系统。该技术通常与其他节能技术混合使用，比如利用 Cache 来缓存写操作从而产生 I/O 突发周期[5]，或将请求重定向到其他磁盘上以延长该磁盘的空闲周期。

4.2 减少消耗能量的磁盘个数

在由多磁盘组成的存储系统中，减少消耗能量的磁盘个数是降低系统能耗最有效的方法之一，如 PARAID，MAID 和 PDC 等技术。它们通过利用 I/O 访问的负载特性，在负载比较轻的时候关闭一部分磁盘来节能，并通过请求重定向技术或数据再分配技术响应用户请求。另外一种方法是利用用户访问的热度特性将热点数据聚集到小部分的磁盘上，进而关闭其它磁盘来达到节能的目的，这种方法的典型代表如 PDC 技术。

5 典型能耗技术讨论与总结

典型的能耗解决方案主要从单个磁盘、磁盘阵列和存储系统这 3 个层次上探讨节能技术。下面将分别分析和讨论这 3 个层次的典型能耗解决方案。

5.1 单个磁盘的能耗研究

5.1.1 前沿研究成果介绍和讨论

截至目前，针对单个磁盘能耗的研究成果主要有如下 3 个，分别是密歇根大学 H. Huang 等人提出的 FS^2 模型[1]、宾夕法尼亚州立大学 S. Gurumurthi 等人提出的 DRPM 模型[2]和 Intra-Disk Parallelism 模型。

FS^2 模型：

H. Huang 等人为了降低磁头定位延迟对磁盘性能和能耗的影响，于 2005 年提出了 FS^2(FreeSpace Filesystem)文件系统。该文件系统根据用户访问模式动态地复制部分热点数据，并将多个副本保存在文件系统的空闲块内，这样对于每次请求磁头都可自动去访问距其最近的数据副本，从而有效降低了磁头定位延迟时间。这种方法在提高磁盘 I/O 性能的同时还有效地降低了磁盘能耗。实验表明，FS^2 能够减少 41%～68%的磁盘访问时间，同时每次磁盘访问的平均能耗降低 40%～71%。

DRPM 模型：

虽然延长磁盘空闲时间被认为是最有效的磁盘节能技术之一，但 E. V. Carrera 等人[17]的研究结果表明：即使在系统处于轻负载并且有较大缓存的情况下，网络服务器类磁盘的空闲周期仍然极小，此时不能挖掘较多的空闲周期来实现磁盘的节能。由于磁盘启动旋转(Spin Up)和停止旋转(Spin Down)需要较长的时间且消耗较多的能量，因此空闲周期较小时不宜采用磁盘频繁启动旋转/停止旋转的方式进行节能。为了解决上述问题，E. V. Carrera 等人和 S. Gurumurthi 等人提出动态转速(Dynamical RPM，DRPM)模型。动态转速模型基于磁盘在不同转速时所消耗的能量不同这一事实，将磁盘原有固定的同一转速分为多个转速级别，并根据负载变化自动在不同转速之间切换。在 S. Gurumurthi 等人提出的模型中，磁盘最大转速为 12000RPM，最小转速为 3600RPM，转速步长为 600RPM。实验结果表明，多速磁盘模型在系统处于重负载时仍然能够有效地降低能耗。

Intra-Disk Parallelism 模型：

为了进一步挖掘磁盘内的并行性，S. Gurumurthi 等人提出了 Intra-Disk Parallelism 模型，即通过增加物理硬件的方式，将传统磁盘中寻道、旋转、数据访问等串行执行的多个步骤并行执行，从而提高单个磁盘的 I/O 性能。该模型将磁盘的并行性分为 4 类：即 D_k，A_l，S_m，H_n，分别表示磁盘栈级别、磁头臂级别、磁盘面级别和磁头级别的并行。

5.1.2 前沿研究成果总结和分析

FS^2 模型的主要价值是利用磁盘的空闲空间来复制部分热点数据，从而有效降低磁头定位延迟。FS^2 的一个不足之处在于其仅仅关注单个磁盘性能提升和节能，而未考虑磁盘阵列甚至整个磁盘存储系统内的数据冗余策略。单个磁盘的性能提升和能耗降低对海量磁盘存储系统的总体性能和总能耗的作用有限。另外，FS^2 内复杂的副本同步机制必然带来较大的开销，使得该技术并不能简单地移植到其他环境中。

DRPM 模型的主要价值在于其打破磁盘单一转速的模式，将磁盘转速分为多个级别，使得磁盘依据工作负载的变化自动选择合适的转速进行数据存取。DRPM 的不足之处是在真实的磁盘内部实现难度较大，目前仅有少数磁盘厂商推出了两级转速的磁盘，但在实际存储系统内的大规模应用还有待时日。

Intra-Disk Parallelism 模型的主要价值在于其充分挖掘磁盘各部件冗余所带来的并行性，但其不足之处也在于实现难度较大，这一点与 DRPM 相似。

从如上 3 个针对单个磁盘能耗研究的模型来看，后面两种模型利用单个磁盘的工作特点节能需要付出巨大的代价。但是 FS^2 是实际可行的，通过对磁盘上数据和副本的有效布局可有效地降低磁盘的能耗。

5.2 磁盘阵列的能耗研究

5.2.1 前沿研究成果介绍和讨论

由于仅针对单个磁盘的节能技术有其较大的局限性，加之磁盘阵列在存储系统内的广泛应用，因此大多数节能技术的研究是在磁盘阵列这个层次上进行的。截至目前，针对磁盘阵列的研究成果主要有如下 8 个，分别是 D. Li 等人提出的

EERAID模型[13]和eRAID模型[12]、Q. Zhu等人提出的Hibernator模型[4]、C. Weddle等人提出的PARAID模型[10]、Q. Zhu等人提出的PA/PB-LRU[7]、X. Yao等人提出的RIMAC[11]、D. Colarelli等人提出的MAID[6]以及E. Pinheiro等人提出的PDC(Popular Data Concentration)模型[8]。

EERAID模型：

美国内布拉斯加大学林肯分校的D. Li等人充分挖掘磁盘阵列内部的冗余信息，将冗余信息的利用和I/O调度策略、阵列控制器级Cache管理策略等结合起来，采用非易失性缓存作为写回策略的Cache来优化写操作请求，提出了EERAID的高能效磁盘阵列。EERAID针对两种最常见的RAID级别：RAID1和RAID5，分别设计了EERAID1和EERAID5。其中EERAID1采用Window Round-Robin(WRR)和Power and Redundancy-Aware Flush(PRF)两个新策略来实现有效的节能。而EERAID5则采用Transformable Read(TRA)和Power and Redundancy-aware Destag(PRD)两种新策略来实现有效的节能。

eRAID模型：

D. Li等人同时充分利用RAID1的冗余特性来重定向I/O请求，提出了eRAID模型。eRAID通过停止旋转部分或整个冗余组的磁盘来降低能耗，同时将系统性能的降低控制在一个可接受的范围内。仿真实验结果表明，在限定的性能影响范围内，eRAID能够节省多达32%的能量。

Hibernator模型：

美国伊利诺伊大学香槟分校Q. Zhu等人针对数据中心类型的负载特征，提出Hibernator模型。在基于动态转速磁盘模型的基础上，在由多种不同转速的磁盘组成的存储系统里，Hibernator将数据迁移到合适转速的磁盘上从而在保证满足性能要求的前提下达到节能的目的。Hibernator中提出了3种关键技术：(1)两级数据布局，热点数据放在磁盘全速旋转的RAID5阵列上以保证高性能，而不活跃的数据放在低速旋转的RAID5阵列上以节省能量；(2)提出一个理论模型来决定优化磁盘设置；(3)快速磁盘移动，以随机移动的方式在两级阵列间快速移动磁盘达到负载平衡。通过这3项技术的应用，Hibernator最大可节约能耗65%。

PARAID模型：

佛罗里达州立大学C. Weddle等人在传统磁盘阵列的基础上，依据系统负载的轻重变化自动调整组成磁盘阵列的活动成员个数，形成一种可动态变换的多档磁盘阵列组织方法，从而在满足性能需求的前提下实现最大程度的节能。实验结果表明，在由5个磁盘构成的原型系统中，PARAID较传统的磁盘阵列在性能和可靠性方面基本相当，但可以节省34%的能量。PARAID典型的应用场合是Web应用负载，在I/O比较轻的情况下运行在低速档状态下可节约能耗，而负载比较重的时候运行在全速档下可保证整个系统的能耗。

PA/PB-LRU模型：

针对到达不同磁盘的不同的I/O行为，比如不同的请求间隔时间分布等，Q. Zhu等人提出了PA-LRU和PB-LRU来提高存储系统的能效，而PB-LRU是在PA-LRU的基础上提出的。由于PA LRU有很多参数需要动态调整，不便于实际应用，因此PB-LRU将PA-LRU的思想具体化，即针对每个磁盘不同的访问特点，将整个Cache按照每个磁盘的特点进行分割，然后给每个磁盘使用，根据系统的访问负载和能耗需求动态地调整。例如当某个磁盘处于休眠状态时，则给该磁盘分配的Cache空间增大，以尽可能地延长其在休眠状态的时间，达到降低能耗的目的。

MAID模型：

为了提升归档存储系统的能效，MAID模型采用额外的磁盘作为缓存磁盘，将热点访问的数据置于新添加的缓存磁盘上，从而最大限度地减少定向到后端磁盘的I/O数量，从而避免后端磁盘频繁地由低能耗的待机状态切换到高能耗的活动状态。MAID适用于归档存储系统，即磁盘数量比较多的大规模存储系统。

PDC模型：

针对存储系统的数据访问频率的差异性，周期性地将热点数据迁移到少数磁盘上，并将访问频率较低的数据集中于剩下的磁盘上。这样可以使绝大部分的I/O请求被尽可能少的磁盘所处理，使处于待机状态的磁盘个数尽可能多，以有效提高系统的能效。PDC对于访问热度比较强的应用效果比较明显，如Web应用等。但是在节约能耗的同时，PDC对于系统的性能有一定的影响。因为PDC将大多数应用请求都集中到了一小部分磁盘上，这样就使这部分磁盘的I/O负载比较重，延长了每个请求的排队时间，从而增大了系统的I/O响应延迟。

RIMAC模型：

在EERAID5模型的基础上，RIMAC模型将内存缓存和RAID5磁盘阵列控制器中NVRAM两层缓存机制组合起来，分别保存数据块信息和校验块信息，利用RAID5编码的异或校验特点，一方面尽可能多地利用两层缓存中的缓存数据服务上层I/O请求，另一方面尽可能地将发送到待机磁盘上的I/O请求重定向到其他的活动磁盘上，这样一方面提高了系统的性能，同时也提高了系统的能效。

5.2.2 前沿研究成果总结和分析

在不同的应用负载下提高以上磁盘阵列级别能效的方法均有各自的特点和优势。EERAID5和RIMAC充分利用了奇偶校验的特性来进行请求转换，从而提高系统的性能和能效。eRAID侧重于建模分析，用吞吐量和响应时间两个目标函数对能耗策略进行了评测。Hibernator是在多个不同转速的磁盘之间进行数据的重布局，从而最大可能地将更多的磁盘切换到更低速的状态。MAID和PDC从负载特征入手，主要利用文件访问频率的差异性来将负载集中到少量的磁盘上，区别仅在于其实现的方式不同，MAID增加了额外的缓存磁盘，PDC则通过文件的迁移完成了热点数据的集中。PA/PB-LRU通过利用每个磁盘不同的负载特征，对标准LRU进行了修改和完善，使其成为有效的缓存算法。PARAID充分挖掘和利用存储系统内空闲的存储空间，将部分磁盘上的数据复制到其余的磁盘上，从而关闭部分磁盘达到节能的目的。

5.3 系统级能耗研究

5.3.1 前沿研究成果介绍和讨论

Diverted Access[9]通过将存储系统内的冗余数据分开存储，使请求仅定向到部分磁盘上而达到节能的目的。Diverted Access充分利用分布式存储系统中的数据冗余特性(如多副本机制)重定向数据请求，将用户请求聚集在一部分存储节点上，从而可以让剩下的存储节点处于低能耗状态。

Write Offloading[14]通过在多组磁盘阵列组成的存储系统中，将要写到待机磁盘上的部分数据重定向到其他磁盘阵列组中的活动磁盘上，以尽可能地延长磁盘待机时间，并降低磁盘启停的切换频率。对于写请求比较多的应用来说，Write Offloading 技术的节能效果比较明显。

Pergamum[15]则是针对归档存储系统的节能技术。Pergamum 在每个节点添加一定量的 NVRAM 来存储数据签名、元数据以及其他一些较小规模的数据项，从而使延迟写、元数据请求以及磁盘间的数据验证等操作均可以在磁盘处于待机状态的情况下进行。

随着 Flash 存储介质的发展，Flash 和 Disk 混合组成的存储系统日益受到人们的关注。如文献[18-20]就是利用 Flash 来做 Disk 的缓存，进而减少 DRAM 和磁盘消耗的能量，同时大大提高存储系统的性能。Flash 介质以其很高的随机读性能优势弥补了 DRAM 和磁盘之间的性能差距，同时其低能耗的特征也大大降低了系统对 DRAM 的需求，可以让磁盘更长时间地处于低能耗的状态，以达到节能的目的。

5.3.2 *前沿研究成果总结和分析*

系统级的节能技术主要通过优化配置数据、添加诸如 NVRAM 等特殊存储设备来存储某些特定的数据等技术而实现节能。与单个磁盘级以及磁盘阵列级的节能技术相比较而言，系统级的节能方法更多地利用新的软硬件技术，从全局的角度上进行资源的优化配置。因此这是未来节能技术的研究热点之一。

另一个研究热点就是针对特殊应用采用特殊的节能技术，如科学计算、Web 应用(PARAID)等。这些应用的访问特征比较明显，比较容易挖掘，从而可以指导系统级优化来达到节能的目的。随着计算机处理能力的快速发展，采用数据挖掘技术来优化磁盘存储系统的能耗也是一个发展方向。

5.4 磁盘存储系统节能技术总结和讨论

上面分别从单个磁盘、磁盘阵列和存储系统级 3 个层面对磁盘存储系统的能耗进行了说明，表 2 对上述的节能技术进行了对比和总结。

表 2 典型节能技术对比

节能策略	级别	节能途径			
		冗余	多速	缓存	不平衡负载
Intra-Disk Parallelism[3]	磁盘级	是	否	否	否
DRPM[2]	磁盘级	否	是	否	否
FS2[1]	磁盘级	是	否	否	否
EERAID[13]	磁盘阵列级	是	是	是	否
eRAID[12]	磁盘阵列级	是	否	否	否
Hibernator[4]	磁盘阵列级	否	是	否	是/数据重分配
PARAID[10]	磁盘阵列级	是	否	否	是
PA/PB-LRU[7]	磁盘阵列级	否	是	是	否
MAID[6]	磁盘阵列级	否	否	是	是/数据重分配
PDC[8]	磁盘阵列级	否	否	否	是/数据重分配
RIMAC[11]	磁盘阵列级	是	否	是	是/请求重定向
Diverted Access[9]	系统级	是	否	是	是/请求重定向
Write Off-Loading[14]	系统级	否	否	是	否
EXCES[14]	系统级	否	否	是	是/数据重分配
C-Burst[14]	系统级	否	否	是	是/数据重定向
Pergamum[15]	系统级	仅用于归档存储系统			

结束语 磁盘存储系统的能耗已经受到越来越多研究者和生产厂商的关注和重视，而能耗也成为计算机系统设计中的一个重要衡量标准。从嵌入式设备、桌面级计算机、分布式系统到云存储系统，能耗在系统总运行成本中所占的比例越来越大。随着存储技术的发展和进步，磁盘将成为信息存储的主要媒介。原先基于磁带的传统备份/归档系统，也逐渐采用低能耗的磁盘作为备份/归档数据存储的主要存储设备。随着全球能源危机的日益临近，如何有效利用电能，降低环境污染已经成为需要重点研究的问题。

计算机系统结构的变迁是以计算(CPU)为中心到以存储(I/O)为中心的转变。全世界数字信息总量每年以成倍的速度递增。据统计，2001 年全球新增的数字化信息为 60 亿 GB，2002 年为 120 亿 GB，2003 年则达到 240 亿 GB。在这样的背景之下，存储市场得到了空前的发展，存储的研究在整个计算机系统结构中也显得越来越重要，而数据密集型和 I/O 密集型应用问题更是当前高性能计算的主要应用。总体说来，计算机的存储研究呈现出以下的趋势：大容量、网络化、容错性和高效性。因此，可以预见，未来对于分布式高性能计算和存储系统的能耗研究将成为主要方向。

参 考 文 献

[1] Huang Hai, Huang Wanda, Shin G K. FS2: Dynamic Data Replication in Free Disk Space for Improving Disk Performance and Energy Consumption[C]// Brighton, UK. Proceedings of the 20th ACM Symposium on Operating Systems Principles (SOSP). New York, NY, USA: ACM, 2005: 263-276

[2] Gurumurthi S, Sivasubramaniam A, Kandemir M, et al. DRPM: Dynamic Speed Control for Power Management in Server Class Disks[C]// San Diego, CA, USA. Proceedings of the International Symposium on Computer Architecture (ISCA). New York, NY, USA: ACM, 2003: 169-181

[3] Sankar S, Gurumurthi S, Stan R M. Intra-Disk Parallelism: An Idea Whose Time Has Come[C]// Beijing, China. Proceedings of 35th the International Symposium on Computer Architecture (ISCA). New York, NY, USA: ACM, 2003: 303-314

[4] Zhu Qingbo, Chen Zhifeng, Tan Lin, et al. Hibernator: Helping Disk Arrays Sleep through the Winter[C]// Brighton, UK. Proceedings of the 20th ACM Symposium on Operating Systems Principles (SOSP). New York, NY, USA: ACM, 2005: 177-190

[5] Papathanasiou E A, Scott L M. Energy Efficient Prefetching and Caching[C]// Boston, MA, USA. Proceedings of the USENIX 2004 Annual Technical Conference (USENIX). Berkeley, CA, USA: USENIX, 2004: 255-268

[6] Colarellt D, Grunwald D. Massive Arrays of Idle Disks for Storage Archives[C]// Baltimore, MD, USA. Proceedings of the 2002 ACM/IEEE Conference on Supercomputing (ICS). Los Alamitos, CA, USA: IEEE, 2002: 1-11

[7] Zhu Qingbo, Zhou Yuanyuan. Power-Aware Storage Cache Management[J]. IEEE Transaction on Computers: 2005, 54(5): 587-602

[8] Pinheiro E, Bianchini R. Energy Conservation Techniques for Disk Array-Based Servers[C]// Malo, France. Proceedings of the 18th International Conference on Supercomputing (ICS). New York, NY, USA: ACM, 2004: 68-78

[9] Pinheiro E, Bianchini R, Dubnichi C. Exploiting Redundancy to Conserve Energy in Storage Systems[C]// Saint Malo, France. Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS). New York, NY, USA: ACM, 2006: 15-26

(下转第 31 页)

度的决策机制等实际问题研究得还不够，可操作性还较差。

结束语 随着 Internet 应用的日益增多，主体间交易、协作及合作的需求会越来越多，对行为安全也提出了更高的要求。本文所介绍的信任管理和信任评估技术是解决此类问题的一种有效方案，现有的一些信任管理系统和信任评估模型也已做出了有效的尝试，但都还存在一些问题与不足。未来将主要研究如何将信任管理和信任评估有机结合起来，这需要重点解决好以下 3 个问题：(1)规范的信任信息描述机制。在 Internet 这个大型的软件运行平台中，信任信息所蕴含的内容将极其丰富，而且与应用的上下文密切相关，需要采用一种通用的、机器可理解的方式来规范地描述信任信息。(2)动态的信任信息收集、分析与评估机制。Internet 应用的安全性需求是动态变化的，参与主体的信任度也会随着协作或合作次数的增多而动态变化。因此，全面地收集主体的信任信息，动态地对信任信息进行分析，并实时地对主体的信任程度给出评价至关重要。(3)合理的信任形成及决策机制。合理地建立初始信任关系，并动态地调整信任关系，以及基于信任度对协作或合作的风险做出评估，从而形成灵活的、情境相关、适应用户动态需求的决策机制，是开放环境下满足灵活、动态安全性的必然要求。

参 考 文 献

[1] Marti S, Garcia-Molina H. Limited reputation sharing in P2P system[C]// Proceedings of the 5th ACM Conference on Electronic commerce. New York, NY, USA, May 2004

[2] 杨芙清，梅宏，吕建，等. 浅论软件技术发展[J]. 电子学报，2002，30(12A)：1901-1906

[3] 沈昌祥，张焕国，冯登国，等. 信息安全综述[J]. 中国科学 E 辑：信息科学，2007，37(2)：129-150

[4] Garnbetta D. Can we trust trust[M]// Gambetta D, ed. Trust: Making and Breaking Cooperative Relations. Blackwell: Oxford Press, 1990: 213-237

[5] Grandison T, Sloman M. A survey of Trust in Internet Applications[C]// IEEE Communications Surveys and Tutorials, Fourth Quarter 2000. 2001

[6] Blaze M, Feigenbaum J, Lacy J. Decentralized Trust Management[C]// IEEE Conference on Security and Privacy. Oakland, California, USA, 1996

[7] Chu Yang-hua, Feigenbaunn J, LaMaeehia B, et al. REFEREE: Trust management for Web applications [J]. Computer Networks and ISDN systems, 1997, 29(8-13): 953-964

[8] 吕建，马晓星，陶先平，等. 网构软件的研究与进展[J]. 中国科学 E 辑：信息科学，2006，36(10)：1037-1080

[9] Beth T, Borcherding M, Klein B. Valuation of thust in open network[C]// Proceeding of the European Symposium on Research in Security(ESORICS). Brighton: Springer-Verlag, 1994: 3-18

[10] 徐锋，吕建，等. 一个软件服务协同中信任评估模型的设计[J]. 软件学报，2003，14(06)：1043-1051

[11] 唐文，陈钟. 基于模糊集合理论的主观信任管理模型研究[J]. 软件学报，2003，14(8)：1401-1408

[12] Jøsang A. An Algebra for Assessing Trust in Certification Chains[C]// The proceedings of NDSS'99, Network and Distributed System Security Symposium. The Internet Society, San Diego, 1999

(上接第 5 页)

[10] Weddle C, Oldham M, Qian Jin, et al. PARAID: A Gear-Shifting Power-Aware RAID[C]// San Jose, CA, USA. Proceedings of the 5th USENIX Conference on File and Storage Technologies (FAST). Berkeley, CA, USA: USENIX, 2007: 245-260

[11] Yao Xiaoyu, Wang Jun. RIMAC: A Redundancy-based, Hierarchical Cache Architecture for Energy-efficient Storage Systems [C]// Leuven, Belgium. Proceedings of the 2006 EuroSys Conference(EuroSys). New York, NY, USA: ACM, 2006: 249-262

[12] Li Dong, Wang Jun. eRAID: A Queueing Model Based Energy Saving Policy [C]// Monterey, CA, USA. Proceedings of 14th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS). Washington, DC, USA: IEEE Computer Society, 2006: 77-86

[13] Li Dong, Wang Jun. EERAID: Energy-efficient Redundant and Inexpensive Disk Array[C]// Leuven, Belgium. Proceedings of 11th ACM SIGOPS European Workshop. New York, NY, USA: ACM, 2004

[14] Narayanan D, Donnelly A, Rowstron A. Write Off-Loading: Practical Power Management for Enterprise Storage[C]// San Jose, CA, USA. Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST). Berkeley, CA, USA: USENIX, 2008: 253-267

[15] Storer W M, Greenan M K, Miller L E, et al. Pergamum: Replacing Tape with Energy Efficient, Reliable, Disk-Based Archival Storage[C]// San Jose, CA, USA. Proceedings of the 6th USENIX Conference on File and Storage Technologies (FAST). Berkeley, CA, USA: USENIX, 2008: 1-16

[16] Zedlewski J, Sobti S, Garg N, et al. Modeling Hard-Disk Power Consumption[C]// San Francisco, CA, USA. Proceedings of the 1st USENIX Conference on File and Storage Technologies (FAST). Berkeley, CA, USA: USENIX, 2003: 217-230

[17] Carrera V E, Pinheiro E, Bianchini R. Conserving Disk Energy in Network Servers[C]// San Francisco, CA, USA. Proceedings of the 17th International Conference on Supercomputing(SC). New York, NY, USA: ACM, 2003: 86-97

[18] Kgil T, Roberts D, Mudeg T. Improving NAND Flash Based Disk Caches[C]// Beijing, China. Proceedings of the 35th International Symposium on Computer Architecture (ISCA). Washington, DC, USA: IEEE Computer Society, 2008: 327-338

[19] Useche L, Guerra J, Bhadkamkar M, et al. EXCES: External Caching in Energy Saving Storage Systems[C]// Salt Lake City, UT, USA. Proceedings of the 14th IEEE International Symposium on High-Performance Computer Architecture(HPCA). Washington, DC, USA: IEEE Computer Society, 2008: 89-100

[20] Chen Feng, Zhang Xiaodong. Caching for Bursts (C-Burst): Let Hard Disks Sleep Well and Work Energetically[C]// Bangalore, India. Proceedings of 2008 International Symposium on Low Power Electronics and Design(ISLPED). New York, NY, USA: ACM, 2008: 141-146

[21] Son S W, Chen Guangyu, Kandemir T M. Disk Layout Optimization for Reducing Energy Consumption[C]// Cambridge, MA, USA. Proceedings of the 19th International Conference on Supercomputing(ICS). New York, NY, USA: ACM, 2005: 274-283