

云计算环境中能效评估方法^①

韩珂¹, 蔡小波², 容会¹, 向丽萍¹

¹(昆明冶金高等专科学校 电气学院, 昆明 650033)

²(云南农业大学 基础与信息工程学院, 昆明 650201)

摘要: 云时代的到来使得数据中心的高能耗问题愈发严重, 如何实现高性能、低功耗且具有 QoS 保障的高能效云环境已经成为研究热点. 论文对能效的基本概念进行了阐述, 并对能效的处理模式包括计算模型和测量方法以及基准测试进行了归纳总结, 在此基础上对能效的关键技术进行深入的分析, 最后归纳总结了高能效云环境面临的问题和挑战.

关键词: 高性能计算; 云计算; 基准测试; 能效

Energy Efficiency Evaluation in Cloud Computing Environment

HAN Ke¹, CAI Xiao-Bo², RONG Hui¹, XIANG Li-Ping¹

¹(School of Electrical Engineering, Kunming Metallurgy College, Kunming 650033, China)

²(College of Basic Science and Information Engineering, Yunnan Agricultural University, Kunming 650201, China)

Abstract: High energy consumption issue is one of the major obstacles for the new cloud era. The energy efficient cloud environment with high performance, and low power consumption has received extensive attention. In this paper, the definition of energy efficiency is given. The principles, the mathematical expression and the measuring and calculation approaches for implementing high energy efficiency are concluded and summarized. On this basis, the key technology is deep analyzed and studied, at last the problems and challenges are summarized and analyzed.

Key words: high performance computing; cloud computing; benchmark; energy efficiency

云计算是一种基于互联网的计算机新方式, 通过互联网上异构、自治的服务为个人和企业用户提供按需汲取的计算. 随着云计算时代的到来, 更多的计算资源和存储资源集中在云端, 数据中心的高能耗逐渐成为突出的问题. 据谷歌公布的数据显示, 谷歌全球能源消费达到 2.6 亿瓦特, 其中 80%-90% 来源于谷歌的 37 个数据中心, 这相当于一个小型城市的总能耗, 或者说这相当于一座标准核电站 1/4 的输出功率. 高能耗成为云计算系统最为严重的问题之一. 因此如何在为用户提供服务质量(QoS)保障和较高性能的同时, 管理云计算中互联网数据中心(IDC)能耗, 实现高性能、低功耗同时具有 QoS 保障的高能效云系统是目目前云计算研究所面临的一个重大挑战.

对高能效云计算的研究, 首先是要找到一种能效

的度量模型和测量方法, 以及一种基准的评估方法来衡量云系统的能效. 能效指标对云计算存在着多方面的影响. 一方面, 在云计算架构下部署的应用在性能、延迟、正常服务运行时间方面有着严格的要求. 如何在保证服务质量的要求下, 还统筹考虑优化云计算系统的能效性能达到既定的要求是一个关键问题; 另一方面, “云”的服务原则是用户按需获取资源, 在用户对于系统资源需求量同等的前提下, 如何有效地配置云中的资源, 在保证系统整体性能的基础上, 使得系统能效最大, 实现服务提供商的利益最大化也是一个重要问题. 直观地看, 能源和效率的综合可以很好的度量性能和能耗之间的关系, 使两者统一, 目前学术研究和商业应用上通常选取能源和效率进行具体化的度量和运算来综合评价云系统的能效.

① 基金项目: 云南省教育厅科学研究基金(2014Z076, 2014Y495, 2011FZ202)

收稿时间: 2015-06-24; 收到修改稿时间: 2015-08-10 [doi: 10.15888/j.cnki.csa.005026]

1 能效的基本概念、来源和在云环境中的应用

目前针对能效的相关研究主要按照物理学的观点, 指在能源利用中, 发挥作用的与实际消耗的能源量之比. 能效可以通过“能源”和“效率”2种度量来综合评价, 表示为单位时间内系统的“性能”和“耗电量”的比值. 文献[1]将能效定义为由基准测试获得的单位瓦特的性能值.

在云计算环境下, 从云服务提供商的角度看, 能效是对云系统设备能耗是否被有效利用的描述. 了解数据中心设备能耗的利用情况, 有利于针对云环境在硬件及软件方向研究节能方案的设计. 从消费者角度看, 能效是指为终端用户提供的服务与所消耗的总能源量之比, 即能量按比例的计算. 所谓“提高能效”, 是指用更少的能源投入提供同等的能源服务, 或是在同等能源消耗上提供更多的性能比例. 能效已被证明是在降低能耗方面获得成本效益最佳的途径, 同时带来巨大的环境效益, 并对消费者提供了积极回报.

2 云能效的计算模型

2.1 基于性能/能耗的能效模型

云环境下能效可以通过单位时间内系统的“性能”和“耗电量”的比值来描述. 通常用性能除以能耗可以消除单位时间, 因此能效单位最终可以被描述为 J/W . 性能通过运行应用程序或计算任务的执行时间或工作量来计算, 而能耗描述为单位时间内系统消耗的能量, 一般采用简单的功率模型, 大部分现有研究都采用焦耳来衡量电能的消耗. 目前关于云系统的能效模型和计算方法主要存在以下几种:

(1)通过比较制冷系统和其他非-IT 资源的消耗评估整个设施的能源效率, 对于数据中心, 绿色网格定义的 Power Usage Effectiveness(PUE)是数据中心消耗的所有能源与 IT 负载使用的能源之比, 是 Data Center Infrastructure Efficiency(DCIE)的对比^[23].

(2)使用单精度浮点计算次数和瓦特的比值来表示系统能效模型. 超级计算机是云环境下数据中心的能耗大户, 对于超级计算机的能效计算分为两部分, 性能是参数 MFLOPS(百万次浮点计算), 由 Linpack 基准测试软件测得, 而功耗单位由精确功率仪器测得瓦特(Watt)值, 文献[20]以“MFLOPS/Watt”表示其能效值, 评估每秒单位瓦特的浮点操作数(FPLOPS).

(3)以云计算系统的单位耗能所能完成的运算量来

定义能效, 文献[21]用负载单位 (U)来衡量计算任务量的大小, 定义处理器在 1GHz 频率下 1s 完成的运算量, 采用 CPU 频率与使用率的乘积来计算, 并对云环境总体性能进行定量的分析和测量, 从运算和能耗这两个角度来衡量系统的性能, 为能效优化提供了依据.

(4)能效表示为单位能量完成的有效任务, 而如何度量有效任务是一个难题. 文献[22]提出定义一种原子的任务单元, 任务单元可以根据算法设计, 如排序算法, 本文可以定义排序 100 条数据的任务量作为单位 1, 或者在数据库系统中定义一个事务为一个任务单元, 或是一次用户请求和响应作为任务单元.

(5)能效表示为 QoS 与能源消耗的比值, 文献[23]总结了实现绿色云环境下兼顾能源和可靠性的自优化框架所面临的挑战, 提出以 QoS/能耗作为衡量云系统能效的标准. 另外, 还有专门的软件用于评估运行在计算资源上的某个应用或中间件的效率.

(6)使用功率的变化来衡量能效优化的效果. 文献[24]用大量实验说明云计算的能效特性, 提出了一种专注于云计算的高能效模型, 指出能效度量是在固定负载下的计算机能耗, 系统能效中也使用了功率或能量作为能耗单位. 文献[25]分析了云环境下不同应用类型、负载、节点类型还有虚拟机大小都影响了系统能效, 并提出针对不同应用类型的单位能耗的负载 ASEE(Application Specific Energy Efficiency)来计算能效.

2.2 性能与能耗的测量方法

能效模型包含性能和能耗 2 个部分, 目前很多研究提出了性能与能耗的度量模型与测量方法.

(1)数据中心能耗测量

通常能耗取值采用功率作为单位, 大部分现有研究都采用“焦耳”来衡量电能的消耗. 功率是单位时间内能耗的大小 J/S , 即瓦(W). 关于云计算环境下能耗的管理主要涉及系统能耗的监控和测量. Kansal 等人^[36]提出了一个间接的虚拟机能耗测量机制 Joulemeter, 使得虚拟化平台也能像硬件上提供给服务器的功耗测量功能一样使用能耗管理机制. Stoesss 等人^[27]假设每个硬件部件的详细能耗模型是已知的, 提出了虚拟机能耗测量机制. 文献[28]给出了服务器、内存、硬盘、模板以及网卡的能耗模型. 此外, 对于异构子节点的云计算环境来说, 电功率的测量较困难, 一般将其划

分为几个同构的子环境然后再进行计算. 一个典型的机架式服务器的能源消耗^[29]如表 1.

表 1 一台典型的机架式服务器的能源消耗

部件	峰值耗能	数量	耗能总量	能耗百分比
CPU	40 W	2	80 W	37.60%
内存	9 W	4	36 W	16.90%
硬盘	12 W	1	12 W	5.60%
PCI 插槽	25 W	2	50 W	23.50%
主板	25 W	1	25 W	11.70%
风扇	10 W	1	10 W	4.70%
系统总能耗			213 W	

(2)性能测试

通常服务器的性能通过运行应用程序或计算任务的方法来监测, 测试指标一般有两种, 一种是执行时间, 另一种是工作量. 比如将 CPU 整数与浮点数运算、内存的随机/顺序读写速度, 存储 I/O 速度及 WEB 系统应用并发连接数作为服务器各个分部件的性能指标, 因此性能描述可以*/S 为单位.

文献[30]以计算任务的执行时间来度量性能, 文献[31]简单地设定一个任务, 用任务的完成时间来代替任务量来对性能进行度量. 文献[21]在实际测量中, 采用 CPU 频率与使用率的乘积来衡量特定 CPU 完成的负载任务大小. 文献[32]将性能理解为数据运算性能, 分为整数运算和浮点数运算. 而整数运算特性, 一般将编译、压缩、视频压缩转换和 XML 处理等功能作为 CPU 处理器整数运算的性能指标. 目前的 CPU 性能测试项包括编程、压缩、编译、组合优化、人工智能、基因序列搜索、国际象棋人工智能计算、物理学量子计算、视频压缩、离散时间仿真、寻路算法、XML 处理. 针对浮点数运算特性的性能测试项包括: Fluid Dynamics (流体力学)、Quantum Chemistry (量子化学)、Quantum Chromodynamics (量子力学)、Physics/CFD (物理计算流体力学)等.

2.3 云能效的基准测试

目前有很多专门的基准测试标准用于比较不同硬件设施和软件的效率. 包括 SPECpower, 用于评估层次结构服务器(volume server class)和 multi-node calss computers 的电能消耗和性能特点. 文献[37]通过使用尽可能少的能耗来对固定数量的记录进行排序, 以此来评估计算节点能耗和性能的折中关系. 针对分布式环境下 QoS 性能评估问题, 相关研究指出了影响系统性能的度量参数主要有系统的响应时间、吞吐量、可靠性、安全性等^[14]. 尽管用户主要关注的服务质量度量因子是系统的响应时间, 但它不能反应系统的整体

服务质量及体现用户与提供者的约束关系, 具有一定的局限性. 文献[15]提出在云计算环境下达到客户层 SLA 要求的情况下, 如何尽量降低系统能耗, 并以此来进行云资源部署, 使系统整体的功耗最小. 研究并未在真实的实验环境中进行验证, 且研究仅停留在能耗的层次, 并没有探讨能耗最小化下的云计算系统所能提供的运算能力. 文献[16]提出未来解决能源问题最好的办法是增大能效或降低能耗, 作者提出 ACES——自动服务器配置系统, 实现数据中心在达到负载的要求下降低能耗, 以此达到系统提高能效的目的. 文献[21]同样以降低数据中心的能耗为出发点, 提出了一种在线算法, 算法可以动态关闭数据中心中的低负载服务器, 动态地调整数据中心的负载分配, 以此来整体降低数据中心的能源消耗. 文献[5]提出了一种方便计算的能效模型, 并给出了能效最大值的存在性和发生条件证明, 但是以同构子节点为假设条件, 并且也未对系统的 QoS 进行评估, 忽视了用户使用云计算服务所定义的 SLA. 等人^[34]则针对异构系统提出了一种性能和能耗的基准测试方法. SPEC[Standard Performance Evaluation Corporation] 标准性能评估测试评估机构是一个全球性的、权威的第三方应用性能测试组织, 主要确立、修改以及认定一系列服务器应用性能评估的标准. 该组织于 2007 年 12 月发布了 SPECpower_ssj2008, 第一项用于评测系统级服务器与运算性能相关的功耗基准测试工具. 该工具生成的服务器负载可以为服务器端加载通用 Java 商业应用程序, 可以评估服务器系统在 10%负载的待机模式到 100%满负载状况下的功耗数据, 服务器系统包括刀片、工作站和机架式服务器; 同时该工具还可以分布扩展, 以多线程方式在不同的系统操作环境件移植. 该工具可以作为一项指标来比较不同服务器之间能耗和性能, 还可帮助提升服务器效率.

3 提高能效的关键技术

所谓“提高能效”, 是指用更少的能源投入提供同等的甚至更多的能源服务. 要实现高能效的云系统一般采用两种处理模式: 一是在微机和主机层面讨论能耗管理技术如休眠技术^[2]、DVFS^[3]、软件优化^[4]和硬件性能改善^[5]等, 用于减少资源使用量以降低能耗. 二是在数据中心宏观层面, 使用资源管理系统(RMS), 比如绿色资源管理^[6,7]、热管理^[8,9]、工作负载整合^[10,11]、

任务调度^[12-13]等来优化调度和负载均衡策略以减少所需要的物理节点。另外,在云计算环境下虚拟化技术使得两种模式下能效的提高更加明显,如在虚拟机技术^[14,15]、虚拟机迁移^[16-18]等云环境下能效的优化。最终通过分别计算各个能耗部件的能效,相加求得云环境下整个数据中心的能源效率。

3.1 节能技术

云计算的节能机制按照不同的分类标准有不同的分类标准,按照功率管理的方式主要可分为动态功率管理(dynamic power management, DPM)技术和静态功率管理(static power management, SPM)技术两类;按照降低能耗阶段的不同可分为关闭/开启技术(resource hibernation)、动态电压/频率调整(dynamic voltage and frequency scaling, DVFS)技术以及虚拟机(virtualization)技术三类,前者主要降低空闲能耗,后者主要降低任务执行能耗。其中DPM的主要前提是云系统中所面临的负载状况是随时间不断变化的,它允许根据当前的性能需求对功率状态进行动态调整,使用DPM技术的主要策略有DVFS和虚拟化。相比而言,SPM主要利用高效的硬件设备,如CPU、硬盘存储、网络设备、UPS和能源提供设备等,设备的结构改变通常能减少能耗。

3.2 基于SLA约束的能效优化

如何实现高效云中的QoS保障,目前的研究主要使用约束满足条件对虚拟机部署进行建模,文献[39-41]约束条件为用户的服务等级协议,以达到最大化节省能量的目的。基本思想是最大化空闲物理机数,通过关闭空闲物理主机来节省能量。

针对分布式环境下QoS性能评估问题,相关研究指出了影响系统性能的度量参数主要有系统的响应时间、吞吐量、可靠性、安全性等^[14]。尽管用户主要关注的服务质量度量因子是系统的响应时间,但它不能反应系统的整体服务质量及体现用户与提供者的约束关系,具有一定的局限性。文献[15]提出在云计算环境下达到客户层SLA要求的情况下,如何尽量降低系统能耗,并以此来进行云资源部署,使系统整体的功耗最小。研究并未在真实的实验环境中进行验证,且研究仅停留在能耗的层次,并没有探讨能耗最小化下的云计算系统所能提供的运算能力。文献[16]提出未来解决能源问题最好的办法是增大能效或降低能耗,作者提出ACES——自动服务器配置系统,实现数据中

心在达到负载的要求下降低能耗,以此达到系统提高能效的目的。文献[17]同样以降低数据中心的能耗为出发点,提出了一种在线算法,算法可以动态关闭数据中心中的低负载服务器,动态地调整数据中心的负载分配,以此来整体降低数据中心的能源消耗。文献[5]提出了一种方便计算的能效模型,并给出了能效最大值的存在性和发生条件证明,但是以同构子节点为假设条件,并且也未对系统的QoS进行评估,忽视了用户使用云计算服务所定义的SLA。

文献[38]提出一种基于熵优化和动态加权的资源评估模型,筛选出满足用户QoS和系统最大化的资源,实现最优调度,减少能耗,实现负载均衡和提高系统利用率。文献[39]提出在云计算环境下达到客户层SLA要求的情况下,如何尽量降低系统能耗,并以此来进行云资源部署,使系统整体的功耗最小。文献[40]提出未来解决能源问题最好的办法是增大能效或降低能耗,作者提出ACES——自动服务器配置系统,实现数据中心在达到负载的要求下降低能耗,以此达到系统提高能效的目的。文献[41]同样以降低数据中心的能耗为出发点,提出了一种在线算法,算法可以动态关闭数据中心中的低负载服务器,动态地调整数据中心的负载分配,以此来整体降低数据中心的能源消耗。

3.3 基于能效的资源管理/任务调度研究

云环境中计算资源都是通过虚拟的方式提供给用户,对于云模式下虚拟机资源的部署事实上是一个多目标的组合优化问题。已经证明,基于多个不相关可加度量的QoS资源调度策略的选择是NP问题,通常需要采用启发式算法或智能优化算法解决。传统的以最小化任务执行时间的批作业调度启发式算法包括蚁群算法、Max-min、Min-min、遗传算法等。这些算法通常以优化任务的执行时间、负载均衡为目标,使调度系统能够获得好的性能指标和能源消耗。高能效的调度策略包括:应用程序负载感知、自适应经济成本感知、服务器集群的信息包括配置、能源消耗和温度状态感知、热管理系统和温度感知的负载配置、研究应用独立和建立性能模型、计算机网站的负载均衡等^[45]。

博弈论作为一个有效的工具在网络通信的拥塞控制、路由协议、P2P的激励机制设计等方面取得了良好的效果。RanjanPal等人对给定QoS情形下IaaS提供商对价格的博弈,以及QoS与价格的博弈这两种情形

分析, 得到存在唯一的最优纳什均衡^[41]. Ashraf A1 Daoud 等人采用斯坦科利伯格博弈, 对大型数据中心为了最大化自己的效益而进行的基于需求的定价模型进行建模, 分析得到数据中心的最优化定价策略^[43]. Jorge Londono 等人对于云计算中固定的物理资源的应用部署引入了搭配博弈的概念, 但是搭配博弈的纳什均衡是一个 NP 难问题, 作者采取了一种简化的, 实际驱动的搭配博弈, 证明了其纳什均衡的存在性^[44].

4 高效云面临的问题和挑战

首先, 能效指标对云计算存在着多方面的影响. 一方面, 在云计算架构下部署的应用在性能、延迟、正常服务运行时间方面有着严格的要求. 如何在保证服务质量的要求下, 还统筹考虑优化云计算系统的能效性能达到既定的要求; 另一方面, “云”的服务原则是用户按需获取资源, 在用户对于系统资源需求量同等的前提下, 如何有效地配置云中的资源, 在保证系统整体性能的基础上, 使得系统能效最大, 实现服务提供商的利益最大化. 云计算模式利用相对集中的资源提供分布式的服务, 不同的云服务提供商和用户有不同的成本要求和 QoS 需求, 不同的业务类型有不同的性能特征指标. 面对不同的提供商、用户和业务类型, 如何得到高效云环境的有效评估方法, 如何量化云计算系统高效的“程度”?

其次, 云计算系统是一个节点异构、多变量、非线性以及各种不确定干扰因素构成的复杂环境, 难以得到有关能耗、性能、QoS 的精确的数学模型. 现有的研究由于建模的误差、复杂环境的变化, 导致能效值呈现出不确定性. 能耗、性能和 QoS 之间关系复杂, 密切耦合, 相互影响, 相互制约, 能耗的降低往往伴随着性能的损失和服务质量的下降, 如何明确三者之间的折衷关系和影响能效的相关因素?

最后, 虚拟化是云计算的基石, 理论上高度虚拟化的云计算就能产生更少的热量, 进而可以降低系统的能源成本. 事实上, 服务器的整合以及虚拟化往往导致了吞吐量下降以及其他的性能问题, 而性能下降又会导致任务完成所需的能源与时间增加, 这将导致云计算数据中心能效具有不确定性. 此外, 基于对云计算模式下能效与系统 QoS 性能指标之间关系的定量分析, 如何设计合理的资源部署优化机制, 在满足不同 SLA 等级用户 QoS 的条件下实现系统能效的最大化,

基于 QoS 与能效之间的关系, 如何设计出合理的资源调度算法, 最终实现满足不同 SLA 等级用户 QoS 需求和系统最大能效的资源调度策略.

参考文献

- 1 Gaeke BR, Husbands P, Li XS, et al. Memory-intensive benchmarks: IRAM vs. cache-based machines. Proc. International Parallel and Distributed Processing Symposium. IPDPS 2002, Abstracts and CD-ROM. IEEE, 2001.
- 2 Chase JS, Anderson DC, Thakar PN, et al. Managing energy and server resources in hosting centers. ACM SIGOPS Operating Systems Review, ACM, 2001, 35(5): 103-116.
- 3 Snowdon DC, Ruocco S, Heiser G. Power management and dynamic voltage scaling: Myths and facts. 2005.
- 4 Kappiah N, Freeh VW, Lowenthal DK. Just in time dynamic voltage scaling: Exploiting inter-node slack to save energy in MPI programs. Proc. of the 2005 ACM/IEEE Conference on Supercomputing. IEEE Computer Society. 2005. 33.
- 5 EcoInfo 2011. From windows 95 to windows 7. EcoInfo report.
- 6 Carrera EV, Pinheiro E, Bianchini R. Conserving disk energy in network servers. Proc. of the 17th Annual International Conference on Supercomputing. ACM. 2003. 86-97.
- 7 Greenpeace 2011. How dirty is your data? Greenpeace report.
- 8 Sharma RK, Bash CE, Patel CD, et al. Balance of power: Dynamic thermal management for internet data centers. Internet Computing, IEEE, 2005, 9(1): 42-49.
- 9 Merkel A, Bellosa F. Balancing power consumption in multiprocessor systems. ACM SIGOPS Operating Systems Review, 2006, 40(4): 403-414.
- 10 Jung G, Joshi KR, Hiltunen MA, et al. A cost-sensitive adaptation engine for server consolidation of multitier applications. Middleware 2009. Springer Berlin Heidelberg, 2009: 163-183.
- 11 Lefevre L, Orgerie AC. Towards energy aware reservation infrastructure for large-scale experimental distributed systems. Parallel Processing Letters, 2009, 19(3): 419-433.
- 12 Yang CY, Chen JJ, Kuo TW, et al. An approximation scheme for energy-efficient scheduling of real-time tasks in heterogeneous multiprocessor systems. Proc. of the Conference on Design, Automation and Test in Europe.

- European Design and Automation Association, 2009: 694-699.
- 13 Da Costa G, De Assuncao MD, Gelas JP, et al. Multi-facet approach to reduce energy consumption in clouds and grids: the GREEN-NET framework. Proc. of the 1st International Conference on Energy-Efficient Computing and Networking. ACM. 2010. 95-104.
- 14 Brey T, Lamers L. Using virtualization to improve data center efficiency. The Green Grid, Whitepaper. 2009, 19.
- 15 Torres J, Carrera D, Hogan K, et al. Reducing wasted resources to help achieve green data centers. IEEE International Symposium on Parallel and Distributed Processing, 2008. IPDPS 2008. IEEE. 2008. 1-8.
- 16 Voorsluys W, Broberg J, Venugopal S, et al. Cost of virtual machine live migration in clouds: A performance evaluation. Cloud Computing. Springer Berlin Heidelberg, 2009: 254-265.
- 17 Liu J, Zhao F, Liu X, et al. Challenges towards elastic power management in internet data centers. 29th IEEE International Conference on Distributed Computing Systems Workshops, 2009. ICDCS Workshops' 09. IEEE. 2009. 65-72.
- 18 Lefèvre L, Orgerie AC. Designing and evaluating an energy efficient cloud. The Journal of Supercomputing, 2010, 51(3): 352-373.
- 19 Grid G. The Green Grid data center power efficiency metrics: PUE and DCiE. Green Grid report, 2007.
- 20 Feng W C, Scogland T. The green500 list: Year one. IEEE International Symposium on Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE, 2009. 1-7.
- 21 宋杰,李甜甜,闫振兴,等.一种云计算环境下的能效模型和度量方法.软件学报,2012,23(2):200-214.
- 22 宋杰,侯泓颖,王智,等.云计算环境下改进的能效度量模型.浙江大学学报(工学版),2013,1:006.
- 23 Bahsoon R. Green cloud: Towards a framework for dynamic self-optimization of power and dependability requirements in green cloud architectures. In: Babar MA, Gortoni, eds. Proc. of the 4th European Conf. on Software Architecture (ECSA 2010). Copenhagen. 2010. 510-514.
- 24 Orgerie AC, Lefèvre L, Gelas JP. Save watts in your grid: Green strategies for energy-aware framework in large scale distributed systems. 14th IEEE International Conference on Parallel and Distributed Systems, 2008. ICPADS'08. IEEE. 2008. 171-178.
- 25 Metri G, Srinivasaraghavan S, Shi W, et al. Experimental analysis of application specific energy efficiency of data centers with heterogeneous servers. 2012 IEEE 5th International Conference on Cloud Computing (CLOUD). IEEE. 2012. 786-793.
- 26 Kansal A, Zhao F, Liu J, et al. Virtual machine power metering and provisioning. Proc. of the 1st ACM Symposium on Cloud Computing. ACM. 2010. 39-50.
- 27 Stoess J, Lang C, Bellosa F. Energy management for hypervisor-based virtual machines. Proc. of the USENIX Annual Technical Conference. 2007. 1
- 28 孙大为,常桂然,陈东,等.云计算环境中绿色服务级目标的分析,量化,建模及评价.计算机学报,2013,36(7):1509-1525
- 29 Fan X, Weber WD, Barroso LA. Power provisioning for a warehouse-sized computer. ACM SIGARCH Computer Architecture News. ACM, 2007, 35(2): 13-23.
- 30 Kumar K, Lu YH. Cloud computing for mobile users: Can offloading computation save energy? Computer, 2010, 43(4): 51-56.
- 31 Abdelsalam HS, Maly K, Mukkamala R, et al. Analysis of energy efficiency in clouds. Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, 2009. COMPUTATIONWORLD'09. Computation World. IEEE, 2009. 416-421.
- 32 郭兵,沈艳,邵子立.绿色计算的重定义与若干探讨.计算机学报,2009,32(12):2311-2319.
- 33 Rivoire S, Shah MA, Ranganathan P, et al. JouleSort: a balanced energy-efficiency benchmark. Proc. of the 2007 ACM SIGMOD international conference on Management of data. ACM. 2007. 365-376.
- 34 Van HN, Tran FD, Menaud JM. Performance and power management for cloud infrastructures. 2010 IEEE 3rd International Conference on Cloud Computing (CLOUD). IEEE. 2010. 329-336.
- 35 Goudarzi H, Ghasemazar M, Pedram M. Sla-based optimization of power and migration cost in cloud computing. 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid). IEEE,

2012. 172-179.
- 36 Xiong K, Perros H. SLA-based resource allocation in cluster computing systems. Proc. of IEEE International Symposium on Parallel and Distributed Processing (IPDPS 2008). 2008. 1-12.
- 37 左利云,曹志波,董守斌.云计算虚拟资源的熵优化和动态加权评估模型.软件学报,2013,24(8):1937-1946.
- 38 Goudarzi H, Ghasemazar M, Pedram M. SLA-based Optimization of Power and Migration Cost in Cloud Computing. 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid). 2012. 172-179.
- 39 Guenter B, Jain N, Williams C. Managing cost, performance, and reliability tradeoffs for energy-aware server provisioning. 2011 Proc. of IEEE INFOCOM. 2011. 1332-1340.
- 40 Lin M, Wierman A, Andrew LLH, et al. Dynamic right-sizing for power-proportional data centers. 2011 Proc. of IEEE INFOCOM. 2011. 1098-1106.
- 41 Ranjan P, Pan H. On the Economics of Cloud Markets. eprint arXiv: 1103.0045. Mar,2011.
- 42 Daoud AA, Agarwal S, Alpcan T. Brief announcement: cloud computing games: pricing services of large data center. Proc. of the 23rd International Conference on Distributed Computing, 2009.
- 43 Londono J, Bestavros A, Teng SH. Colocation games: and their application to distributed resource management. Proc. of the 2009 Conference on Hot Topics in Cloud Computing, 2009.
- 44 Cao JW, Li KQ, Stojmenovic I. Optimal power allocation and load distribution for multiple heterogeneous multicore server processors across clouds and data centers. IEEE Trans. on Computers, 2014, 63(63): 45-58