

云计算系统中数据中心的节能算法研究*

张小庆¹, 贺忠堂², 李春林¹, 张恒喜^{1,3}, 钱琼芬^{1,4}

(1. 武汉理工大学计算机科学与技术学院, 武汉 430063; 2. 中国科学院计算技术研究所 广东电子工业研究院, 广东 东莞 523808; 3. 徐州空军学院基础部, 江苏 徐州 221000; 4. 空军雷达学院四系, 武汉 430019)

摘要: 简要介绍了云计算的定义和特点, 重点研究了云计算数据中心的高能耗问题, 对目前的节能算法进行了分类, 重点综述了基于 DVFS 的节能算法、基于虚拟化的节能算法以及基于主机关闭/开启的节能算法, 并对算法的优缺点和适用环境作了比较分析。最后总结了云计算数据中心的能耗管理中进一步的研究难题。

关键词: 云计算; 数据中心; 节能环保; 动态电压/频率调整; 虚拟机

中图分类号: TP301.6 **文献标志码:** A **文章编号:** 1001-3695(2013)04-0961-04

doi:10.3969/j.issn.1001-3695.2013.04.001

Research on energy saving algorithm of datacenter in cloud computing system

ZHANG Xiao-qing¹, HE Zhong-tang², LI Chun-lin¹, ZHANG Heng-xi^{1,3}, QIAN Qiong-fen^{1,4}

(1. College of Computer Science & Technology, Wuhan University of Technology, Wuhan 430063, China; 2. Guangdong Electronics Industry Institute, Institute of Computing Technology, Chinese Academy of Sciences, Dongguan Guangdong 523808, China; 3. Dept. of Fundamental, Xuzhou Air Force Academy, Xuzhou Jiangsu 221000, China; 4. Dept. of No. 4, Air Force Radar Academy, Wuhan 430019, China)

Abstract: This paper briefly described the definition and characteristics of cloud computing, and focused the high energy consumption problem of cloud datacenter. Based on the classification of the energy saving algorithm, it researched the three kind of energy saving algorithms in a big way, including the energy saving algorithm based-DVFS, energy saving algorithm based-virtualization and the energy saving algorithm based-turn off/on of hosts, and made a comparative analysis of the advantages, disadvantages and applicable scene of these algorithms. Finally, it summed up the further research problems for energy consumption management in cloud computing datacenter.

Key words: cloud computing; datacenter; energy saving and environmental protection; DVFS; virtual machine

0 引言

云计算^[1,2]作为最新兴的计算模式为用户即用即付的行为方式提供了动态灵活的架构和 QoS 服务, 可以将其定义为一种基于 Internet 的计算方式。通过这种方式, 将基于 Internet 的相关 IT 能力以服务的方式提供给用户, 用户无须了解提供云服务的细节, 也不必具有相应的专业知识以及设备操作能力, 通过 Internet 获取需要的服务, 整个运行方式如同电网^[3]。云计算拥有超大规模、虚拟化、数据存储可靠、高度共享、可扩充以及廉价等特点^[4]。目前, 许多大型 IT 厂商都推出了各自的云计算平台, Google App Engine^[5]、Amazon EC2^[6]、IBM Blue Cloud^[7]及 Microsoft Azure^[8]等云计算基础设施的应用都较为广泛。而云计算领域的研究主要集中在海量数据存储、资源管理、虚拟化、任务调度、数据中心节能算法以及云安全等方面。本文将重点研究云系统数据中心的能耗问题及节能算法。

能耗问题已经成为制约云计算发展的关键问题, EPA 的报告指出, 2006 年美国总能源水泵的 1.5% 被用在数据中心, 并且这种趋势将日趋严重。2007 年 4 月的报道指出, 信息与通信技术领域带来了全球 CO₂ 排放量的 2%, 与航空业持平;

同时研究表明, CO₂ 排放量保持 15% ~ 30% 的下降才能保证 2020 年前全球气温升高低于 2℃。2010 年, Google 共排放了 146 万公吨的 CO₂, 100 条搜索的用电量相当于 60 W 灯泡持续亮 28 min。可以看出, 解决云系统数据中心的高能耗问题已经成为一个环境问题。

1 数据中心节能算法分类

目前, 云计算的节能机制按照不同的分类标准有不同的分类。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)^[9] 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。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 技术的主要有 DVFS 策略和虚拟化策略。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

收稿日期: 2012-06-30; **修回日期:** 2012-09-21 **基金项目:** 国家自然科学基金资助项目 (60970064, 60773211, 61171075); 国家教育部新世纪优秀人才计划资助项目 (NCET-08-0806); 软件开发环境国家重点实验室开放课题 (SKLSDE-2009KF-2-02); 霍英东教育基金资助项目 (121067)

作者简介: 张小庆 (1984-), 男, 博士研究生, 主要研究方向为云计算 (zxqzty@whut.edu.cn); 贺忠堂, 男, 博士 (后), 主要研究方向为云计算及产业化; 李春林, 女, 教授, 博导, 主要研究方向为网络计算; 张恒喜, 男, 副教授, 博士研究生, 主要研究方向为云计算; 钱琼芬, 博士研究生, 主要研究方向为云计算。

的结构改变通常能减少能耗。

2 基于 DVFS 的节能算法

DVFS 是使用在现代处理器上控制 CPU 功耗的有效节能方式之一,其主要思想是:当 CPU 未被完全利用时,通过降低 CPU 的供电电压和时钟频率主动降低 CPU 性能,这样可以带来立方数量级的动态功耗降低,并且对性能不会产生影响。例如考虑某系统仅有一个任务执行,负载要求 10 个周期完成,而任务的截止时间是 100 s,这种情况下处理器可能降速至 1/10 周期/s,满足了截止时间同时节省了能耗。这种动态的频率降低过程比起以最大速率运行任务然后在剩余时间内闲置 CPU 更加有效。DVFS 已在包括移动终端设备以及服务器中的 CPU 得到了广泛使用。

文献[10]在实时云服务环境中提出三种基于 DVFS 的能量感知的虚拟机提供策略。云服务模型定义为 $\{T_i(r_i, c_i, d_i, p_i, f_i) | i=1, 2, \dots, n\}$ 。其中: n 为任务数; r_i 表示任务 i 的开始时间; c_i 表示最差状况的执行时间; d_i 表示相对截止时间; p_i 表示周期; f_i 表示完成时间。子任务 T_i 开始于时间 r_i , 且要求最差执行时间为 c_i , 为完成应用目标, T_i 应在 $r_i + d_i$ 内完成; p_i 给出了 T_i 的周期, 使得应用在 $r_i + kp_i$ 时再次开启一个执行时间 c_i 的子任务, 且应在 $r_i + kp_i + d_i$ ($k=0, 1, \dots$) 时完成。虚拟机模型定义为 $\{V_i(u_i, m_i, d_i) | i=1, 2, \dots, n\}$ 。其中: u_i 表示应用请求的 CPU 利用率, m_i 表示虚拟机基准 MIPS, d_i 表示截止时间。用户提交服务到虚拟机后, 资源提供者利用 DVFS 提供虚拟机以减少能耗, 以下是提出的三种基于 DVFS 的节能策略:

a) Lowest-DVFS。调整 CPU 速率至最低, 满足虚拟机截止时间为准, 每个虚拟机以请求的 MIPS 执行任务, 该策略在任务到达率较低、能够接收所有请求的情况下能耗最少。

b) δ -advanced-DVFS。为克服 Lowest-DVFS 策略较低的服务接收率问题, 该策略为当前的虚拟机请求的 MIPS 提高 $\delta\%$ 。处理器的伸缩 s 由下式决定, 对于给定的虚拟机请求集中的任务 T_k , 在时间 t 下, 每个虚拟机的比例共享正比于 $w_i/(d_i - t)$ 。

$$s = \min \left\{ 1, \left(1 + \frac{\delta}{100} \right) \times \frac{1}{Q_k} \sum_{i \in T_k} \frac{w_i}{d_i - t} \right\}$$

其中, $\delta\%$ 根据系统负载预先进行定义。

c) Adaptive-DVFS。当请求到达率及它们的服务时间提前预知时, 可以分析得到最优化的伸缩模型。考虑一个到达率 λ 和服务率 μ 的 M/M/1 队列模型, 若处理器的速率伸缩设为 s , 那么根据队列模型可知, 平均响应时间 $RT = 1/s\mu - \lambda$, 响应时间 RT 应小于等于平均截止时间 d , 以满足实时服务请求 $1/s\mu - \lambda \leq d$ 。因此, 最优化伸缩 s^* 为

$$s^* = \frac{1}{\mu} \left(\lambda + \frac{1}{d} \right)$$

Adaptive-DVFS 管理服务请求持续时间 h 的平均到达速率 $\hat{\lambda}$, 平均服务率 $\hat{\mu}$ 和平均截止时间 $\hat{d}$, 它在时间 t 及给定请求集 T_k 时根据下式调整处理器速率。

$$s = \max \left\{ \min \left\{ 1, \frac{1}{\hat{\mu}} \left(\hat{\lambda} + \frac{1}{\hat{d}} \right) \right\}, \frac{1}{Q_k} \sum_{i \in T_k} \frac{w_i}{d_i - t} \right\}$$

文献实验结果表明 δ -advanced-DVFS 在单位功耗内的收益表现方面最好。由于 CPU 性能可以根据系统负载自动调整, 而 Adaptive-DVFS 的性能则受到具体队列模型的限制。

文献[11]基于 DVFS 技术提出了一种启发式调度算法, 用

来降低并行任务在集群环境中执行产生的能耗。该算法针对并行任务图中非关键路径上的任务, 在不影响整个任务完成时间的前提下, 降低非关键任务所调度 CPU 的电压来降低能耗。文献[12]也进行了类似的研究, 提出了集群环境中的基于能量感知的任务调度算法, 提出通过控制适当的电压级别来减少功耗。文献[13]针对嵌入式多处理器系统, 提出了基于能耗感知的启发式任务调度算法 EGMS 和 EGMSIV。算法同时考虑了任务调度顺序和电压的动态调整, 并提出利用能耗梯度作为任务调度的评价指标, EGMSIV 在 EGMS 的基础上实现了同一任务执行时的电压动态调整。

文献[14]研究了具有 DVS 能力的多处理器平台中周期性抢占式硬件实时任务的能量最小化问题, 该文采用分段调度机制, 为每个任务分配一个静态优先级, 一旦任务分配给处理器, 将启动处理器速率分配机制, 减少能耗, 同时保证灵活性。

DVFS 的目的是降低执行能耗。执行能耗定义为: 任务在计算机上运行时, 指令和数据驱动计算机硬件运转所产生的能耗。同一个任务在执行过程中, 其执行功率会随着运行阶段、执行特征的变化而变化^[15]。由于 CMOS 电路动态功率中功率与电压与频率成正比, 降低 CPU 的电压或频率可以降低 CPU 的执行功率。该方法的缺点在于降低 CPU 的电压或频率之后, CPU 的性能也会随之降低^[13]。

3 基于虚拟化的节能算法

另一种改善资源利用率从而降低功耗的手段是虚拟化技术。虚拟化技术是云计算中的关键技术之一, 它允许在一个主机上创建多个虚拟机, 因此减少了硬件资源的使用数量, 改进了资源利用率。虚拟化不仅可以使得共享相同计算节点的应用之间实现性能的隔离, 而且还可以利用动态或离线迁移技术实现虚拟机从一个节点至另一个节点之间的迁移, 而实时的虚拟机重分配可以实现动态的负载合并, 这样虚拟机可以合并到较小数量的物理节点上, 从而转换闲置节点为节能模式。

云数据中心以虚拟机为单位为用户分配资源, 用户通过运行虚拟机执行服务。虚拟机的部署和迁移是云数据中心虚拟机能量优化管理的重要问题, 图 1 是虚拟机的迁移过程。

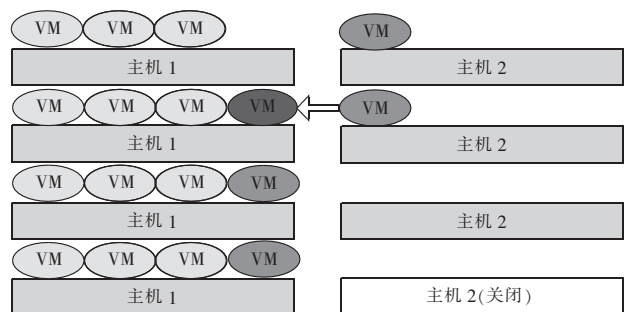


图 1 基于虚拟机迁移的节能方法

文献[16]将能耗管理技术与虚拟化技术结合起来, 为大规模数据中心开发了一种能耗优化管理方法 VirtualPower。该方法支持虚拟机独立运行自身的能耗控制方法, 并能合理协调不同虚拟化平台之间以及同一个虚拟化平台上不同虚拟机之间的能耗控制请求, 实现对能耗的整体优化。

文献[17]使用约束满足问题^[18]对虚拟机部署进行建模, 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。实现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。

虚拟机部署问题可以考虑为一个全局虚拟机分配矩阵 $R_g = [R_{1,1}, \dots, R_{i,j}, \dots, R_{m,s}]$, $R_{i,j}$ 表示虚拟机 j 分配到应用程序 i 的数量;物理机到虚拟机的映射矩阵即虚拟机的部署由矩阵 P 表示, $P_{h,v} \in \{0,1\}$, 当值为 1 时表示虚拟机 v 分配到物理机 h 上。部署解决方案必须满足如下约束条件:

$$\forall h \in \{1, 2, \dots, n\} \quad \sum_{v=1}^l P_{h,v} \times \text{CPU}(V_v) \leq \text{CPU}(H_h)$$

$$\forall h \in \{1, 2, \dots, n\} \quad \sum_{v=1}^l P_{h,v} \times \text{mem}(V_v) \leq \text{mem}(H_h)$$

即分配到各个虚拟机上的 CPU 和内存总和要在物理机提供的 CPU 和内存的能量范围之内。目标函数是最大化空闲的物理机数 N_{idle} , 即

$$\begin{aligned} \max \quad & N_{\text{idle}} = \sum_{h=1}^n x_h \\ \text{s. t.} \quad & x_h = \begin{cases} 1 & \text{if } \sum_{v=1}^l P_{h,v} = 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

虚拟机部署问题是一个多背包问题的实例,使用约束满足问题可以很容易地引入额外的应用约束,如搭配约束和同类约束等。该策略从约束满足方面考虑虚拟机的部署,在满足用户服务等级协议的前提下,达到最小能源花费的目的。但是,该策略要求资源本身的利用率问题,且当问题域过大时,不能保证得到最优方案。

文献[19]通过将云计算环境中的虚拟资源分配问题形式化为一个路径构建问题,提出了一种高能效的分配策略 EE-VRAS,策略使用受限精华策略的蚁群系统生成优化的资源分配方案,降低了服务器的使用数量及系统能耗。

以上策略均是从减小服务器的使用数量为目标。文献[20]提出将虚拟机的动态迁移^[21]与关闭空闲计算节点相结合的方式,提高物理资源的利用率,平衡电量和性能间的需求,该文提出了动态虚拟机再分配机制,即先将资源的利用率排序,将资源利用率较小的物理主机上的虚拟机以最小电量增加原则迁移出去,将空闲的物理主机关闭,节省电量。文献给出的迁移标准是设定固定的主机资源占用率,虚拟机迁移的目标是使资源占用率保持在占用率的上限和下限值之间。迁移虚拟机的选择算法为 HPG 策略,其基本思想是:当上限违例时,HPG 迁移的虚拟机为相对于虚拟机参数中定义的 CPU 能力拥有最低 CPU 占用率的虚拟机集合,定义为

$$R = \begin{cases} \{S | S \in P(V_j), u_j - \sum_{v \in S} u_v(v) < T_u, \sum_{v \in S} \frac{u_v(v)}{u_j(v)} \rightarrow \min\} & \text{if } u_j > T_u \\ V_j & \text{if } u_j > T_l \\ \emptyset & \text{otherwise} \end{cases}$$

其中: $u_v(v)$ 为虚拟机 v 初始请求的 CPU 能力,定义为虚拟机参数。

文中提出的门限值虽然可以控制主机的 CPU 占用情况,但最优的门限无法通过分析得出,文中通过多次实验分析可以得到门限的近似最优值,且文中没有讨论虚拟机迁移的代价问题^[22]。

文献[23]开发了一种分层的能耗控制系统,包含宿主机和用户级子系统。前者根据用户请求对硬件资源进行合理分配,以使每个虚拟机的能耗不超过规定上限;后者在虚拟机层重新对虚拟硬件资源进行分配,使每个用户任务产生的能耗不超过规定上限。

虚拟化技术提高了物理资源的利用率,降低了执行能耗,但其缺点在于增加了系统管理的复杂性。由于引入虚拟化技

术后,资源管理将涉及虚拟机提供(virtual machine provision)和虚拟机部署(virtual machine placement)两个方面^[24],前者主要负责以虚拟机的形式向应用任务分配资源,该阶段一般与托管应用的 SLA 相关,且提供机制由性能目标所驱动,性能目标选择一般为平均的任务响应时间,单位时间完成作业量来描述;后者主要完成虚拟机到物理主机间的映射,虚拟机的部署问题由与资源管理代价相关的数据中心的策略所驱动。

4 基于主机关闭/开启的节能算法

基于主机关闭/开启技术的节能策略可以分为随机式策略、超时式策略和预测式策略^[25]三类。随机策略将服务器的关闭/开启时机视为一个随机优化模型,利用随机决策模型设计控制算法。超时式策略预先设置一系列超时阈值,若持续空闲时间超过阈值,就将服务器切换到关闭模式,同时,阈值可以固定不变,也可以随系统负载自适应调整。预测式策略在初始阶段就对本次空闲时间进行预测,一旦预测值足够大,就直接切换到关闭模式。三类策略的目标均是最大限度地降低空闲能耗,而缺点在于当计算机启动较长时,会导致性能一定程度的降低。这类策略的主要研究有:

文献[25]提出,由于计算机系统业务请求具有自相似性,导致基于关闭/开启技术的最优节能策略为超时策略;并提出了当空闲时间长度服务 Pareto 分布时,基于截尾均值法小样本情况下, Pareto 分布形状参数的稳健有效估计算法和基于窗口大小自适应技术非平稳业务请求下的 DPM 控制算法。

文献[26]引入负载感知机制,提出了一种在虚拟化计算平台中的动态节能算法,利用动态提供机制关闭不必要的主机子系统来实现节能。与随机策略和预定义的超时策略相比,该方法在保证 QoS 前提下可提高能耗效率。

基于主机关闭/开启的节能技术可与虚拟化中的虚拟机迁移方法结合起来使用,当可以预知负载信息时,该方式可以极大节约空闲主机的闲时能耗。

5 其他节能算法

冷却系统约占云数据中心总能耗的 40%,计算资源的高速运行导致设备温度升高,温度过高不仅会降低数据系统的可靠性,而且会减少设备的使用寿命。因此,必须对云数据中心的冷却设备降温,有效地减少冷却的能量,对云数据中心的稳定运行和节省电量都有重要的意义。

文献[27]提出将数据中心冷却系统考虑在内,在服务器中安装变速风扇和温度传感器,根据服务器的温度调整风扇的转速,既保证了安全又节省电量。

文献[28]提出数据中心的指令数据流包括温度传感器数据、服务器指令、数据中心空间的空调单元数据,该文对这些数据流进行分析,提出了简单灵活的模型,根据给定的负载分布和冷却系统配置可以预测数据中心热分布,然后静态或手动配置热负载管理系统。

文献[29]提出了数据中心级的电量和热点管理解决方案,PTM(power and thermal management)引擎决定活动服务器的数目的位置,同时调节提供的冷却温度,提高了数据中心的能效。

文献[30]提出了控制数据中心风扇的精细方式,每个机架上的风扇根据其自身的热系统、硬件使用率等信息调整风扇

的速率。

文献[31]综合考虑空间大小、机架和风扇的摆放以及空气的流动方向等因素,提出一种多层次的数据中心冷却设备设计思路,并对空气流和热交换进行建模与仿真,为数据中心布局提供理论支持。

此外,数据中心建成以后,可采用动态冷却策略降低能耗,如对于处于休眠的服务器,可以适当关闭一些制冷设施或改变冷气的走向,以节约成本^[32]。文献[33]针对云数据中心内部热量分配不均衡的问题,首次提出以无线多媒体传感器网络(WMSNs)实时监测局部热点,并利用任务迁移等方法降低热点区域的热负载,以“热点发现—热点定位—特征提取—热点消除”为主线,实现平衡热量分配、提高制冷效率的目的。

除此之外,动态组件钝化(dynamic component deactivation, DCD)^[27]也是在硬件上使用的重要节能手段。

DPM 利用实时的资源使用状况和应用负载状况对能耗实行优化,不足之处在于若负载一直处于峰值状态,功耗并不能减少。

云数据中心节能算法的比较如表1所示。

表1 云数据中心节能算法的比较

文献/策略	虚拟化	目标系统	系统资源	优化目标	节能策略
Lowest-DVFS ^[10]	√	实时系统	CPU	执行能耗	DVFS, DPM
δ-advanced-DVFS ^[10]	√	实时系统	CPU	执行能耗	DVFS, DPM
Adaptive-DVFS ^[10]	√	实时系统	CPU	执行能耗	DVFS, DPM
文献[11, 12]	×	集群系统	CPU	执行能耗	DVFS, DPM
EGMS ^[13]	×	多处理器系统	CPU	执行能耗	DVFS, DPM
EGMSIV ^[13]	×	多处理器系统	CPU	执行能耗	DVFS, DPM
文献[14]	×	实时系统	CPU	执行能耗	DVFS, DPM
VirtualPower ^[16]	√	服务器系统	CPU	执行能耗	虚拟化, DPM
CSP ^[17]	√	服务器系统	CPU, mem	执行能耗	虚拟化, DPM
EEVRAS ^[19]	√	服务器系统	CPU	执行能耗	虚拟化, DPM
文献[20]	√	服务器系统	CPU	执行能耗	虚拟化, 动态迁移, DPM
文献[23]	√	服务器系统	CPU	执行能耗	虚拟化, DPM
文献[25]	×	集群系统	CPU	空闲能耗	关闭/开启, DPM
文献[26]	√	服务器系统	CPU	空闲能耗	关闭/开启, DPM
文献[27~31, 33]	-	-	设备	设备能耗	SPM
文献[27]	-	-	设备	设备能耗	DCD

6 算法应用场景分析

不同的节能策略通常拥有不同的应用场景。基于 DVFS 的节能算法的主要思想是通过动态调整 CPU 的电压和频率使其在不同阶段拥有不同的功率/性能,用不同的功率/性能处理不同的负载类型或不同计算量的任务,在降低执行能耗的同时保证执行性能。虚拟化节能算法实现了计算机资源从物理实体到虚拟实体的过渡,提高了计算资源的使用率。然而,虚拟化节能算法本身要付出高昂的效能代价,因为虚拟化技术通过对底层硬件到高层服务应用的层层虚拟,每一级的虚拟都不可避免地会造成效能的损耗。关闭/开启节能算法通常是针对服务器的关闭/开启时机进行设定或预测,但对于包含大量类型的计算资源的云计算系统而言,如何根据单位时间到达的任务量决定要关闭的服务器数量,以及关闭哪些服务器等问题,都给关闭/开启节能策略带来了难题^[34]。

7 结束语

本文主要研究了云计算系统数据中心的节能算法,并且对目前使用的典型算法进行了分类及比较分析,可以看出,基于

不同的应用环境,每种算法都有其自身的优缺点。综合现有的研究,在数据中心能耗管理方面仍然有进一步的研究重点:a)如何在给定的真实云计算系统中,根据任务类型、到达率及分布决策物理主机的运行状态,并结合 DVFS 和主机关闭/开启技术对系统进行能效优化;b)云是面向服务的,这必定要求满足一定的 QoS 需求,如何定义一种 QoS/能效模型来度量云系统的能耗优化目标,并明确它们之间的主从关系,也是在未来需要进一步关注的问题。

参考文献:

- [1] WEISS A. Computing in the clouds[J]. *Networker*, 2007, 11(4): 16-25.
- [2] 刘鹏. 云计算[M]. 2版. 北京:电子工业出版社, 2011.
- [3] <http://zh.wikipedia.org> [EB/OL].
- [4] GARG S K, BUYYA R. Green cloud computing and environmental sustainability[EB/OL]. <http://www.buyya.com/papers/Cloud-Env-Sustainability2011.pdf>.
- [5] Google app engine[EB/OL]. <http://appengine.google.com>.
- [6] Amazon elastic compute cloud (EC2) [EB/OL]. <http://www.amazon.com/ec2>.
- [7] IBM blue cloud[EB/OL]. <https://www.ibm.com/developerworks/cloud/>.
- [8] Microsoft Azure[EB/OL]. <http://www.microsoft.com/azure>.
- [9] MAGKLIS G, SEMERARO G, ALBONESI D H, et al. Dynamic frequency and voltage scaling for a multiple-clock-domain microprocessor[J]. *IEEE Micro*, 2003, 23(6): 62-68.
- [10] KIM K H, BELOGLAZOV A, BUYYA R. Power-aware provisioning of virtual machines for real-time cloud services[J]. *Concurrency and Computation: Practice and Experience*, 2011, 23(13): 1491-1505.
- [11] WANG Li-zhe, Von LASZEWSKI G, DAYAL J, et al. Towards energy aware scheduling for precedence constrained parallel tasks in a cluster with DVFS[C]//Proc of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing. Washington DC: IEEE Computer Society, 2010: 368-377.
- [12] Von LASZEWSKI G, WANG Li-zhe, YOUNGE A J, et al. Power-aware scheduling of virtual machines in DVFS-enabled clusters[C]//Proc of IEEE International Conference on Cluster Computing and Workshops. 2009: 1-10.
- [13] GOH L K, VEERAVALLI B, VISWANATHAN S. Design of fast and efficient energy-aware gradient-based scheduling algorithms heterogeneous embedded multiprocessor systems[J]. *IEEE Trans on Parallel and Distributed Systems*, 2009, 20(1): 1-12.
- [14] ALENAWY T A, AYDIN H. Energy-aware task allocation for rate monotonic scheduling[C]//Proc of the 11th IEEE Real Time on Embedded Technology and Applications Symposium. Washington DC: IEEE Computer Society, 2005: 213-223.
- [15] BLUME H, Von LIVONIUS J, ROTENBERG L, et al. OpenMP-based parallelization on an MPCore multiprocessor platform: a performance and power analysis[J]. *Journal of System Architecture*, 2008, 54(11): 1019-1029.
- [16] NATHUJI R, SCHWAN K. VirtualPower: coordinated power management in virtualized enterprise systems[C]//Proc of the 21st ACM SIGOPS Symposium on Operating Systems Principles. New York: ACM Press, 2007: 265-278.
- [17] Van NGUYEN H, TRAN F D, MENAUD J M. Performance and power management for cloud infrastructures[C]//Proc of the 3rd International Conference on Cloud Computing. 2010: 329-336.
- [18] 董存祥, 王文俊, 杨鹏. 基于约束满足问题的应急决策[J]. *计算机工程*, 2010, 36(7): 276-278. (下转第 970 页)

- [11] 陈贵海,李振华. 对等网络:结构、应用与设计[M]. 北京:清华大学出版社,2007:35-40.
- [12] P2P-next[EB/OL]. <http://www.p2p-next.org/>.
- [13] SENTINELLI A, ANSELMO T, FRAGNETO P, *et al.* Packetizing scalable streams in heterogeneous peer-to-peer networks[C]//Proc of IEEE International Conference on Multimedia and Expo. 2011:1-6.
- [14] SHEU T L, WANG Y H. Dynamic layer adjustments for SVC segments in P2P streaming networks [C]//Proc of International Computer Symposium. 2010:793-798.
- [15] LEE T C, LIU Pin-chuan, SHYU W L, *et al.* Live video streaming using P2P and SVC [C]//Proc of the 11th IEEE International Conference on Management of Multimedia and Mobile Networks and Services. 2008:104-113.
- [16] ABBOD O, ZINNER T, PUSSEP K, *et al.* On the impact of quality adaptation in SVC-based P2P video-on-demand systems [C]//Proc of the 2nd Annual ACM Conference on Multimedia System. New York: ACM Press, 2011:223-232.
- [17] LIU Zheng-ye, SHEN Yan-ming, ROSS K W, *et al.* Layer P2P: using layered video chunks in P2P live streaming [J]. *IEEE Trans on Multimedia*, 2009, 11(7):1340-1352.
- [18] SONG Jun-ping, ZHOU Xu, ZHANG Yan, *et al.* A playback length changeable segmentation algorithm for SVC based P2P streaming systems [C]//Proc of International Conference on Communications. 2012:2032-2037.
- [19] ASIOLI S, RAMZAN N, IZQUIERDO E. Efficient scalable video streaming over P2P network [C]//Proc of International UCMedia Conference. 2010:153-160.
- [20] DING Yan, LIU Jiang-chuan, WANG Dan, *et al.* Peer-to-peer video-on-demand with scalable video coding[J]. *Computer Communication*, 2010, 33(14):1589-1597.
- [21] LIU Zheng-ye, SHEN Yan-ming, PANWAR S S, *et al.* Using layered video to provide incentives in P2P live streaming [C]//Proc of Workshop on Peer-to-Peer Streaming and IPTV. New York: ACM Press, 2007:311-316.
- [22] HU Hao, GUO Yang, LIU Yong. Peer-to-peer streaming of layered video: efficiency, fairness and incentive[J]. *IEEE Trans on Circuits and Systems for Video Technology*, 2011, 21(8):1013-1026.
- [23] ZHANG Gui, YUAN Chun. Self-adaptive peer-to-peer streaming for heterogeneous networks using scalable video coding [C]//Proc of the 12th IEEE International Conference on Communication Technology. 2010:1390-1393.
- [24] MUBASHAR M, AHMED T. Smooth video delivery for SVC based media streaming over P2P networks [C]//Proc of IEEE Industry Applications Society Annual Meeting. 2008:447-451.
- [25] ZHANG Meng, XIONG Yong-qiang, ZHANG Qian, *et al.* On the optimal scheduling for media streaming in data-driven overlay networks [C]//Proc of Global Telecommunications Conference. 2006.
- [26] XIAO Xin, SHI Yuan-chun, GAO Yuan, *et al.* LayerP2P: a new data scheduling approach for layered streaming in heterogeneous networks [C]//Proc of IEEE INFOCOM. 2009:603-611.
- [27] COHEN B. Incentives build robustness in BitTorrent [C]//Proc of the 1st Workshop on Economics of Peer-to-Peer Systems. 2003.
- [28] CHU Yang-hua, CHUANG J, ZHANG Hui. A case for taxation in peer-to-peer streaming broadcast [C]//Proc of ACM SIGCOMM Workshop Practice Theory Incentives Networked Systems. New York: ACM Press, 2004:205-212.
- [29] HSU C H, HEFEEDA M. Achieving viewing time scalability in mobile video streaming using scalable video coding [C]//Proc of the 1st Annual ACM Conference on Multimedia System. New York: ACM Press, 2010:111-122.
- [30] CRANLEY N, PERRY P, MURPHY L. User perception of adapting video quality [J]. *International Journal of Human-Computer Studies*, 2006, 64(8):637-647.
- [31] ZINNER T, HOHLFELD O, ABBOD O, *et al.* Impact of frame rate and resolution on objective QoE metrics [C]//Proc of the 2nd International Workshop on Quality of Multimedia Experience. 2010:29-34.
- [32] MONTEIRO J M, NUNES M. A subjective quality estimation tool for the evaluation of video communication systems [C]//Proc of the 12th IEEE Symposium on Computers and Communications. 2007:75-80.
- [33] SINGH K D, KSENTINI A, MARIENVAL B. Quality of experience measurement tool for SVC video coding [C]//Proc of IEEE International Conference on Communications. 2011:1-5.
- (上接第964页)
- [19] 曾智斌, 许力. 云计算中高性价比的虚拟资源分配策略[J]. *计算机系统应用*, 2011, 20(12):55-60.
- [20] BELOGLAZOV A, BUYYA R. Energy efficient allocation of virtual machines in cloud data centers [C]//Proc of the 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing. Washington DC: IEEE Computer Society, 2010:577-578.
- [21] 刘鹏程. 云计算中虚拟机动态迁移的研究[D]. 上海: 复旦大学, 2009.
- [22] VOORSLUYS W, BROBERG J, VENUGOPAL S, *et al.* Cost of virtual machine live migration in clouds: a performance evaluation [C]//Proc of the 1st International Conference on Cloud Computing. Berlin: Springer-Verlag, 2009:254-265.
- [23] STOESS J, LANG C, BELLOSA F. Energy management for hypervisor-based virtual machines [C]//Proc of USENIX Annual Technical Conference. 2007:1-14.
- [24] Van HIEN N, TRAN F D, MENAUD J M. Autonomic virtual resource management for service hosting platforms [C]//Proc of ICSE Workshop on Software Engineering Challenges of Cloud Computing. Washington DC: IEEE Computer Society, 2009:1-8.
- [25] 吴琦, 熊光泽. 非平衡自相似业务下自适应动态功耗管理[J]. *软件学报*, 2005, 16(8):1499-1505.
- [26] RODERO I, JARAMILLO J, QUIROZ A, *et al.* Energy-efficient application-aware online provisioning for virtualized clouds and data centers [C]//Proc of International Conference on Green Computing. Washington DC: IEEE Computer Society, 2010:31-45.
- [27] MEISNER D, GOLD B T, WENISCH T F. PowerNap: eliminating server idle power [J]. *ACM SIGPLAN Notices*, 2009, 44(3):205-216.
- [28] MOORE J, CHASE J S, RANGANATHAN P. Weatherman: automated, online and predictive thermal mapping and management for data centers [C]//Proc of IEEE International Conference on Autonomic Computing. Washington DC: IEEE Computer Society, 2006:155-164.
- [29] PAKBAZNIA E, CHASEMAZAR M, PEDRAM M. Temperature-aware dynamic resource provisioning in a power-optimized datacenter [C]//Proc of Conference on Design, Automation & Test in Europe. 2010:124-129.
- [30] TOLIA N, WANG Zhi-kui, MARWAH M, *et al.* Delivering energy proportionality with non energy-proportional systems: optimizing the ensemble [C]//Proc of Conference on Power Aware Computing and Systems. Berkeley: USENIX Association, 2008:2.
- [31] SAMADIANI E, JOSHI Y, MISTREE F. The thermal design of a next generation data center: a conceptual exposition [C]//Proc of International Conference on Thermal Issues in Emerging Technologies: Theory and Application. 2007:93-102.
- [32] CHEN Gong, HE Wen-bo, LIU Jie, *et al.* Energy-aware server provisioning and load dispatching for connection-intensive internet services [C]//Proc of the 5th USENIX Symposium on Networked Systems Design and Implementation. Berkeley: USENIX Association, 2008:337-350.
- [33] 刘航. WMSNs 在云计算中心节能减排中的关键技术研究[D]. 大连: 大连理工大学, 2011.
- [34] 谭一鸣, 曾国荪, 王伟. 随机任务在云计算平台中能耗的优化管理方法[J]. *软件学报*, 2012, 23(2):266-278.