

融合双速率和工作休眠的虚拟机调度策略及参数优化

金顺福^{1,2,3}, 郝闪闪^{1,2,3}, 王宝帅^{1,2,3}

- (1. 燕山大学信息科学与工程学院, 河北 秦皇岛 066004;
2. 河北省计算机虚拟技术与系统集成重点实验室, 河北 秦皇岛 066004;
3. 河北省软件工程重点实验室, 河北 秦皇岛 066004)

摘要: 受日趋严格的环境标准的影响, 高能耗和高污染成为制约云数据中心发展的重要因素。在保证云数据中心服务质量的前提下, 应用动态电压调整技术, 引入同步多重工作休眠模式, 提出一种新型的虚拟机调度策略。通过构建具有自适应服务率和多重工作休眠的二维连续时间马尔可夫 (Markov) 随机模型, 运用矩阵几何解方法, 从系统节能水平和用户请求平均延迟等方面评估虚拟机调度策略的性能。综合理论分析结果和仿真统计结果, 验证虚拟机调度策略的有效性。从经济学角度出发, 建立系统效用函数, 改进萤火虫智能优化算法, 给出工作休眠参数的优化方案, 实现系统响应性能和节能效果之间的合理平衡。

关键词: 云数据中心; 虚拟机调度策略; 双速率; 多重工作休假排队; 效用函数; 智能优化

中图分类号: TP393

文献标识码: A

Virtual machine scheduling strategy based on dual-speed and work vacation mode and its parameter optimization

JIN Shun-fu^{1,2,3}, HAO Shan-shan^{1,2,3}, WANG Bao-shuai^{1,2,3}

- (1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China;
2. Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004, China;
3. Key Laboratory for Software Engineering of Hebei Province, Qinhuangdao 066004, China)

Abstract: Due to the increasingly strict environmental standards, high pollution and high energy consumption have become the significant factors restricting the development of cloud data centers (CDC). Under the premise of guaranteeing the quality of service (QoS) of CDC, dynamic power management (DPM) technology was applied, synchronous multiple working sleep mode was introduced, and a novel virtual machine (VM) scheduling strategy was proposed. By establishing a two-dimensional continuous-time Markov stochastic model with adaptive service rate and synchronous multiple work vacations, and using the method of a matrix geometric solution, the performance of the VM scheduling strategy was evaluated in terms of energy saving level and average delay of requests. Numerical results with analysis and simulation verify the energy saving effectiveness of the VM scheduling strategy. In order to achieve a reasonable balance between the response performance and the energy-saving effect, a system utility function was constructed from the perspective of economics and design a researching algorithm of the sleep parameter based on the firefly algorithm (FA).

Key words: cloud data centers, virtual machine scheduling strategy, dual-speed, multiple working vacations queue, utility function, intelligent optimization

1 引言

随着互联网服务的广泛普及和云计算技术的

飞速发展, 云数据中心的数量和规模均呈指数级高速增长。国际数据中心 (IDC) 预测 2017 年全球大型云数据中心的数量将会达到 860 万个^[1]。云数据

收稿日期: 2017-08-08; 修回日期: 2017-11-23

基金项目: 国家自然科学基金资助项目 (No.61472342); 河北省自然科学基金资助项目 (No.F2017203141)

Foundation Items: The National Natural Science Foundation of China (No.61472342), The Natural Science Foundation of Hebei Province (No. F2017203141)

中心的耗能是指服务器、空调系统、网络通信设备等各种用能设备所消耗的能源总和。据自然资源保护协会 (NRDC) 报道, 2006 年美国数据中心总耗能量为 610 亿千瓦时, 2013 年总耗能量为 910 亿千瓦时, 顺应这一趋势, 2020 年总耗能量预计达到 1400 亿千瓦时。快速增长的能源成本和日益严格的环境标准使能耗成为运营和维护云数据中心的最大成本所在, 严重制约着云数据中心的发展。因此, 绿色节能技术成为近年来学术界和产业界关注的热点话题, 建立绿色云数据中心是云计算发展的必然趋势。

研究表明, 服务器中最耗能的部件之一是 CPU, 服务器的耗能 CPU 的利用率近似呈线性关系^[2]。不合理的服务器调度策略可能会造成所有的用户请求 (包括对响应性能要求不高的用户请求) 都在高配置服务器中执行, 由此浪费的能耗称为奢侈能耗。合理调整服务器的运行速度, 提高 CPU 的利用率是实现绿色节能最直接的方法。Li^[3]研究了动态电压管理技术 (DPM, dynamic power management), 根据系统的负载情况合理地调整服务器中 CPU 的运行速度。该文献的研究表明, 给定单位时间的能耗水平, 通常存在一个能使系统平均响应时间最小的速度配置方案。Wang 等^[4]提出了一个基于贝叶斯分类器的在线负载预测器和一种基于强化学习的自适应 DPM 技术。根据对未来短时间内系统负载的预测结果, 使用 DPM 技术调整服务器运行速度, 实现能耗的有效节约。以上文献专注于降低系统奢侈能耗的相关技术的研究。

云数据中心服务器的利用率普遍较低, 尤其在夜间和凌晨, 利用率仅为 5%~25%^[5]。所以, 另一种绿色节能的方法是减少空闲能耗。Liao 等^[6]提出了一种节能策略, 将系统服务器分为 3 组, 其中一组服务器始终保持开启状态, 另外 2 组备用服务器则根据系统当前的负载情况动态开启/关闭。Shen 等^[7]提出了一种通过虚拟化技术设计数据中心规模的节能方案, 允许将若干个用户请求合并成一个用户请求以占用较少的服务器, 关闭处于空闲状态的服务器, 达到节能的目的。马晓彤等^[8]在认知无线网络中的基站端引入半休眠模式, 提出一种节能策略。在休眠参数的制约下, 基站根据数据分组的到达情况, 在唤醒阶段、休眠阶段、低速传输阶段和监听阶段之间转换。以上文献专注于降低系统空

闲能耗的相关方案的研究。

在保证云数据中心服务质量的前提下, 以最大限度降低奢侈能耗和空闲能耗为目的, 将 DPM 技术与同步多重工作休眠模式相结合, 本文提出一种新型的融合双速率和工作休眠的虚拟机调度策略。考虑部分虚拟机运行速度的切换及剩余虚拟机工作休眠与正常运行的状态切换, 建立一个具有自适应服务率和多重工作休眠机制的二维连续时间马尔可夫随机模型。利用矩阵几何解方法, 得出系统节能水平、虚拟机利用率和用户请求平均延迟等指标的表达式, 结合理论与仿真实验验证虚拟机调度策略的有效性。为了寻找云数据中心节能效果和响应性能的平衡点, 建立系统效用函数, 给出工作休眠参数的优化算法。

2 能量节省策略及系统模型

2.1 能量节省策略

在传统的云数据中心, 除了处理用户请求需产生的必要能耗外, 运行过程中还存在巨大的能耗浪费^[9,10]。为了最大限度地降低云数据中心的奢侈能耗和空闲能耗, 应用 DPM 技术, 引入同步多重工作休眠模式, 提出一种新型的融合双速率和工作休眠的虚拟机调度策略。

将云数据中心中所有的虚拟机分为主模块和备用模块 2 个部分。主模块中的虚拟机时刻保持开启状态, 其运行速度取决于当前的系统负载情况。备用模块中的虚拟机可能处于工作休眠状态或高速运行状态, 由工作休眠定时器和当前的系统负载情况共同决定。具体来说, 当系统负载较轻时, 令主模块中所有的虚拟机低速运行以降低系统的奢侈能耗, 令备用模块中的虚拟机同时进入随机长度的工作休眠以降低系统的空闲能耗。当系统负载较重时, 为了保证系统的响应性能, 不仅主模块中的虚拟机以高速运行, 备用模块中的虚拟机在本次工作休眠结束后也恢复至高速运行状态。

主模块和备用模块中虚拟机数量分别记为 n 和 m 。为了避免主模块中虚拟机运行速度的频繁切换, 采用双阈值 θ_1 和 θ_2 ($0 < \theta_2 < \theta_1$) 共同控制主模块中虚拟机的状态。当系统中的用户请求数大于阈值 θ_1 时, 主模块中所有的虚拟机以高速运行; 当系统中的用户请求数小于阈值 θ_2 时, 主模块中所有的虚拟机以低速运行。为了保证当主模块中的虚拟机低速运行时, 备用模块中的虚拟机一定处于工作休眠状态,

令 $m \leq n - \theta_2$ 。

为了控制备用模块中的虚拟机状态，引入一个激活阈值 $\theta_3 (\theta_3 > 0)$ 。当系统中总的用户请求数减少到 n 时，备用模块中未结束服务的用户请求立即迁移到主模块中空闲的虚拟机中，服务速率不变。然后，备用模块中的虚拟机同时进入工作休眠状态，启动一个工作休眠定时器，以超低的速率服务之后到达备用模块的用户请求。当休眠定时器到期，即本次工作休眠结束时，如果系统缓存中用户请求的数量超过阈值 θ_3 ，备用模块中的虚拟机立即切换至高速运行状态；否则，备用模块中的虚拟机进入下一次工作休眠，并重启一个新的工作休眠定时器。

云数据中心虚拟机的状态转换过程如图 1 所示。

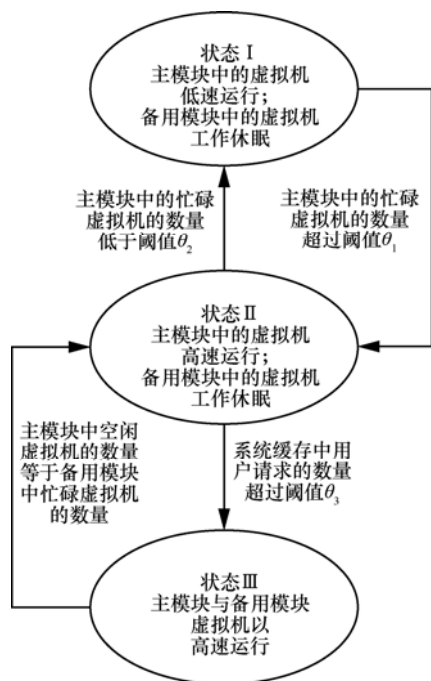


图 1 虚拟机状态的转换过程

状态 I 系统处于该状态时，主模块中的虚拟机以低速运行，备用模块中的虚拟机始终处于工作休眠状态。所有到达系统的用户请求均可以立即占用主模块中空闲的虚拟机，接受低速服务。随着用户请求的不断到达，主模块中越来越多的虚拟机可能被占用，处于忙碌状态。当主模块中忙碌虚拟机的数量超过阈值 θ_1 时，主模块中虚拟机的运行速度从低速切换至高速，即系统从状态 I 转换到状态 II。未结束服务的用户请求将继续占用当前的虚拟机，但服务速率上升。

状态 II 系统处于该状态时，主模块中的虚拟机以高速运行，备用模块中的虚拟机始终处于工作休眠状态。当用户请求到达系统时，如果主模块或备用模块中存在空闲的虚拟机，用户请求将立即占用空闲的虚拟机，接受高速或超低速服务，否则，需要在系统缓存中排队等待。一方面，随着用户请求的到达，系统缓存中用户请求的数量可能越来越多。当工作休眠定时器到期时，如果该数量大于阈值 θ_3 ，备用模块中的虚拟机立即切换至高速运行状态，即系统从状态 II 转换到状态 III。备用模块中未结束服务的用户请求将继续占用当前的虚拟机，但服务速率上升。另一方面，随着结束服务的用户请求的离开，主模块中忙碌虚拟机的数量可能越来越少。当该数量低于阈值 θ_2 时，主模块中虚拟机的运行速度从高速切换至低速，即系统从状态 II 转换到状态 I。未结束服务的用户请求将继续占用当前虚拟机，但服务速率下降。

状态 III 系统处于该状态时，主模块和备用模块中的虚拟机均以高速运行。当用户请求到达系统时，如果主模块或备用模块中存在空闲的虚拟机，用户请求将立即占用空闲的虚拟机，接受高速服务，否则需要在系统缓存中排队等待。随着结束服务的用户请求的离开，主模块中空闲虚拟机的数量可能越来越多，备用模块中忙碌虚拟机的数量可能越来越少。当主模块中空闲虚拟机的数量等于备用模块中忙碌虚拟机的数量时，备用模块中未结束服务的用户请求立即迁移到主模块中空闲的虚拟机中，服务速率不变。然后，备用模块中的虚拟机同时进入工作休眠状态，启动一个工作休眠定时器，即系统从状态 III 转换到状态 II。

2.2 系统模型

假设用户请求到达间隔和工作休眠定时器的长度分别服从参数为 $\lambda (\lambda > 0)$ 和 $\delta (\delta > 0)$ 的指数分布。当虚拟机处于工作休眠状态、低速运行状态和高速运行状态时，假设服务一个用户请求的时间分别服从参数为 $\mu_s (0 < \mu_s < 1)$ 、 $\mu_l (\mu_s < \mu_l < 1)$ 和 $\mu_h (\mu_l < \mu_h < 1)$ 的指数分布。

令随机变量 $N(t) = i, i = 0, 1, 2, L$ 表示 t 时刻系统中用户请求的数量，称为系统水平。令随机变量 $C(t) = j, j = 1, 2, 3$ 分别表示虚拟机处于状态 I、状态 II 和状态 III。 $\{(N(t), C(t)), t \geq 0\}$ 构成一个二维连

续时间马尔可夫链，其状态空间 Ω 表示为

$$\Omega = \{(i, j) \mid i = 0, 1, 2, L, j = 1, 2, 3\}$$

令 $\pi_{i,j}$ 表示稳态下系统水平为 i 且虚拟机处于状态 j 时的概率分布， $\pi_{i,j}$ 定义为

$$\pi_{i,j} = \lim_{t \rightarrow \infty} P\{N(t) = i, C(t) = j\},$$

$$i = 0, 1, 2, L, j = 1, 2, 3$$

令 $\pi_i (i \geq 0)$ 表示稳态下系统水平为 i 的概率向量，则

$$\pi_i = \begin{cases} (\pi_{i,1}, \pi_{i,2}), & i = 0, 1, 2, L, n \\ (\pi_{i,2}, \pi_{i,3}), & i = n + 1, n + 2, n + 3, L \end{cases}$$

该马尔可夫链的稳态概率分布 π 由 $\pi_i (i \geq 0)$ 构成， π 表示为

$$\pi = (\pi_0, \pi_1, \pi_2, L)$$

3 系统模型的稳态分析

在融合双速率和工作休眠的虚拟机调度策略中，不同的系统水平对应不同的虚拟机状态，如表 1 所示。

表 1 系统水平与虚拟机状态的关系

系统水平	初始的虚拟机状态	一步转移后的虚拟机状态
$[0, \theta_2 - 1]$	状态 I	状态 I
θ_2	状态 I	状态 I
	状态 II	状态 I 或状态 II
$[\theta_2 + 1, \theta_1 - 1]$	状态 I	状态 I
	状态 II	状态 II
θ_1	状态 I	状态 I 或状态 II
	状态 II	状态 II
$[\theta_1 + 1, n]$	状态 II	状态 II
	状态 III	状态 II 或状态 III
$n + 1$	状态 II	状态 II
	状态 III	状态 II 或状态 III
$[n + 2, n + m + \theta_3]$	状态 II	状态 II
	状态 III	状态 III
$[n + m + \theta_3 + 1, \infty)$	状态 II	状态 II 或状态 III
	状态 III	状态 III

基于表 1，给出马尔可夫链 $\{(N(t), C(t)), t \geq 0\}$ 的状态转移率如图 2 所示。

令 Q 表示马尔可夫链 $\{(N(t), C(t)), t \geq 0\}$ 的一步转移率矩阵。由表 1 可知，每个系统水平最多对

应 2 个不同的虚拟机状态，故 Q 可划分为若干个具有 2×2 结构的子阵，子阵由图 2 所示的相关状态之间的转移率确定。令 $Q_{u,v}$ 表示经过一步转移后，系统水平为从 $u (u = 0, 1, 2, L)$ 到 $v (v = 0, 1, 2, L)$ 的转移率子阵。为表述方便，将 $Q_{u,u-1}$ 、 $Q_{u,u}$ 和 $Q_{u,u+1}$ 分别记为 B_u 、 A_u 和 C_u 。

当 $u = 0$ 时， A_0 中的元素表示 (0,1) 到 (0,1)、(0,1) 到 (0,2)、(0,2) 到 (0,1) 和 (0,2) 到 (0,2) 的状态转移率。 C_0 中的元素表示 (0,1) 到 (1,1)、(0,1) 到 (1,2)、(0,2) 到 (1,1) 和 (0,2) 到 (1,2) 的状态转移率。

当 $1 \leq u \leq n - 1$ 时， B_u 中的元素表示 $(u, 1)$ 到 $(u - 1, 1)$ 、 $(u, 1)$ 到 $(u - 1, 2)$ 、 $(u, 2)$ 到 $(u - 1, 1)$ 和 $(u, 2)$ 到 $(u - 1, 2)$ 的状态转移率。 A_u 中的元素表示 $(u, 1)$ 到 $(u, 1)$ 、 $(u, 1)$ 到 $(u, 2)$ 、 $(u, 2)$ 到 $(u, 1)$ 和 $(u, 2)$ 到 $(u, 2)$ 的状态转移率。 C_u 中的元素表示 $(u, 1)$ 到 $(u + 1, 1)$ 、 $(u, 1)$ 到 $(u + 1, 2)$ 、 $(u, 2)$ 到 $(u + 1, 1)$ 和 $(u, 2)$ 到 $(u + 1, 2)$ 的状态转移率。

当 $u = n$ 时， B_n 中的元素表示 $(n, 1)$ 到 $(n - 1, 1)$ 、 $(n, 1)$ 到 $(n - 1, 2)$ 、 $(n, 2)$ 到 $(n - 1, 1)$ 和 $(n, 2)$ 到 $(n - 1, 2)$ 的状态转移率。 A_n 中的元素表示 $(n, 1)$ 到 $(n, 1)$ 、 $(n, 1)$ 到 $(n, 2)$ 、 $(n, 2)$ 到 $(n, 1)$ 和 $(n, 2)$ 到 $(n, 2)$ 的状态转移率。 C_n 中的元素表示 $(n, 1)$ 到 $(n + 1, 2)$ 、 $(n, 1)$ 到 $(n + 1, 3)$ 、 $(n, 2)$ 到 $(n + 1, 2)$ 和 $(n, 2)$ 到 $(n + 1, 3)$ 的状态转移率。

当 $u = n + 1$ 时， B_{n+1} 中的元素表示 $(n + 1, 2)$ 到 $(n, 1)$ 、 $(n + 1, 2)$ 到 $(n, 2)$ 、 $(n + 1, 3)$ 到 $(n, 1)$ 和 $(n + 1, 3)$ 到 $(n, 2)$ 的状态转移率。 A_{n+1} 中的元素表示 $(n + 1, 2)$ 到 $(n + 1, 2)$ 、 $(n + 1, 2)$ 到 $(n + 1, 3)$ 、 $(n + 1, 3)$ 到 $(n + 1, 2)$ 和 $(n + 1, 3)$ 到 $(n + 1, 3)$ 的状态转移率。 C_{n+1} 中的元素表示 $(n + 1, 2)$ 到 $(n + 2, 2)$ 、 $(n + 1, 2)$ 到 $(n + 2, 3)$ 、 $(n + 1, 3)$ 到 $(n + 2, 2)$ 和 $(n + 1, 3)$ 到 $(n + 2, 3)$ 的状态转移率。

当 $n + 2 \leq u < \infty$ 时， B_u 中的元素表示 $(u, 2)$ 到 $(u - 1, 2)$ 、 $(u, 2)$ 到 $(u - 1, 3)$ 、 $(u, 3)$ 到 $(u - 1, 2)$ 和 $(u, 3)$ 到 $(u - 1, 3)$ 的状态转移率。 A_u 中的元素表示 $(u, 2)$ 到 $(u, 2)$ 、 $(u, 2)$ 到 $(u, 3)$ 、 $(u, 3)$ 到 $(u, 2)$ 和 $(u, 3)$ 到 $(u, 3)$ 的状态转移率。 C_u 中的元素表示 $(u, 2)$ 到 $(u + 1, 2)$ 、 $(u, 2)$ 到 $(u + 1, 3)$ 、 $(u, 3)$ 到 $(u + 1, 2)$ 和 $(u, 3)$ 到 $(u + 1, 3)$ 的状态转移率。

进一步分析可知，当 $u > n + m + \theta_3 + 1$ 时， $B_u = B_{n+m+\theta_3+1}$ 、 $A_u = A_{n+m+\theta_3+1}$ 及 $C_u = C_{n+m+\theta_3+1}$ 。 Q 可表示为分块三对角形式

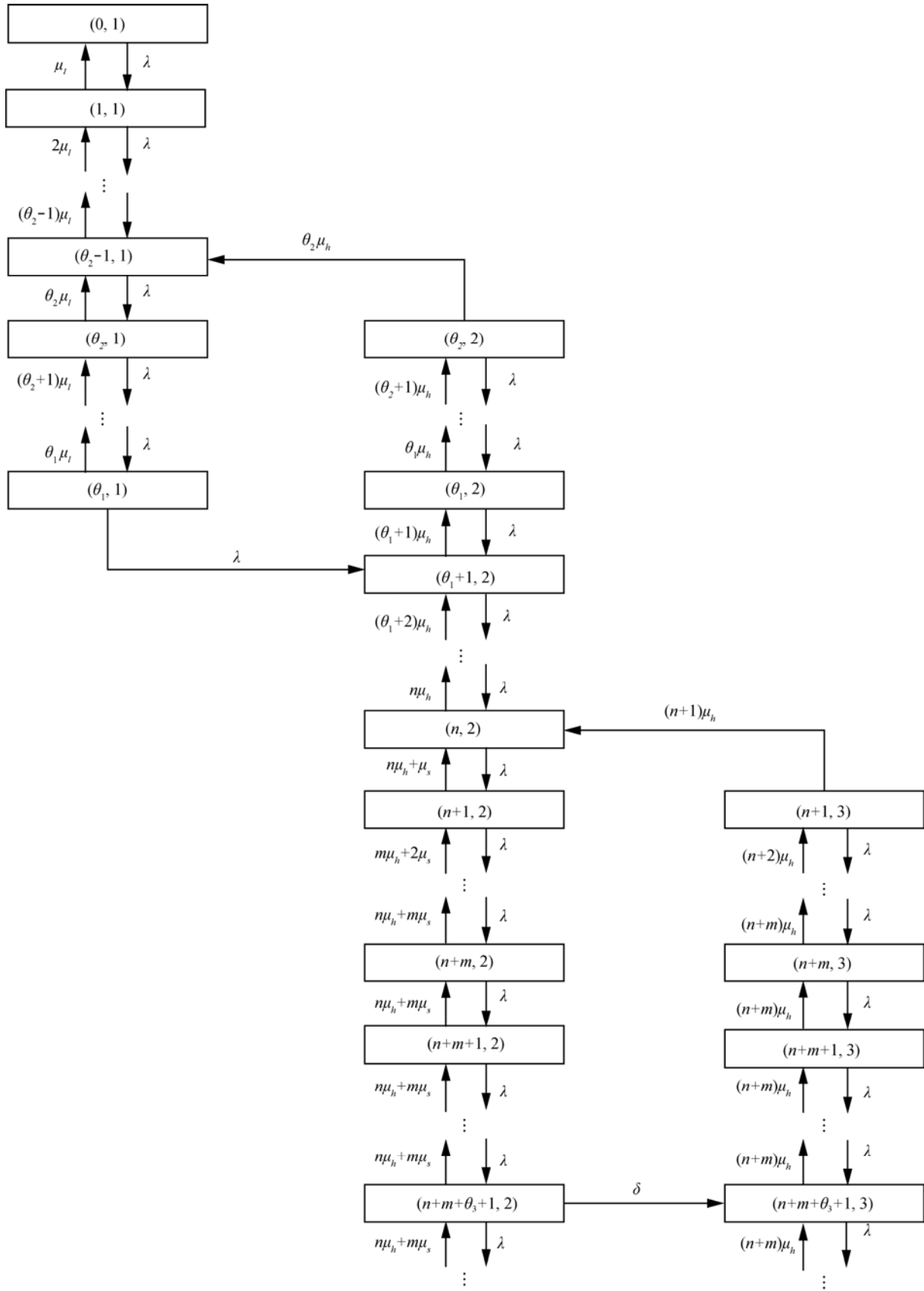


图 2 马尔可夫链的状态转移率

$$S_2 = \sum_{i=n+m+\theta_3+1}^{\infty} \pi_{i,2} \delta m (J_b + J_c) + \pi_{\theta_3,1} \lambda n J_a \quad (8)$$

其中, J_a 、 J_b 和 J_c 分别表示虚拟机由低速状态切换到高速状态、由休眠状态切换到高速状态及休眠边界处监听负载情况的能耗水平。

综合式(7)和式(8), 可得融合双速率和工作休眠的虚拟机调度策略系统节能水平 S 为

$$S = S_1 - S_2 \quad (9)$$

假设虚拟机的能耗水平与虚拟机的运行速度成正比, 令 $k(k > 0)$ 为能耗因子。以 $J_s = k\mu_s$, $J_l = k\mu_l$, $J_h = k\mu_h$, $J_a = 3k(\mu_h - \mu_l)$, $J_b = 3k(\mu_h - \mu_s)$, $J_c = 0.3k\mu_s$ 为例进行系统实验。

4.2 系统实验

为了进一步研究工作休眠参数 δ , 工作休眠服务率 μ_s 和激活阈值 θ_3 对虚拟机调度策略的影响, 进行数值及仿真实验。数值实验在具有高效计算功能和强大显示功能的 Matlab R2011a 环境中基于式(6)~式(9)实现; 仿真实验用 Java 语言在 MyEclipse 2014 环境中实现。Java 语言虽运行稍慢, 但其面向对象的特点可以支持通过类的定义表示事物的多种属性, 仿真实验中定义了具有未到达、等待、运行、结束等属性的用户请求类及具有低速运行、高速运行、空闲、工作休眠等属性的虚拟机类。模拟 10 000 个用户请求的随机到达和服务过程, 仿真实验的最终结果是 10 次模拟结果的平均值。

将本文所提出的虚拟机调度策略与文献[3]中只具有速率调整机制的 DPM 策略进行系统性能的对比实验。在图 3 和图 4 中, 基本虚拟机调度策略的实验结果用实线表示, 基于 DPM 策略的实验结果用虚线表示, 系统仿真结果用“*”表示。

实验中系统参数的设定如表 2 所示。

表 2 系统参数

参数	值
主模块中虚拟机的数量 n	30
备用模块中虚拟机的数量 m	20
双阈值 θ_1 和 θ_2	20, 10
用户请求的到达率 λ	15
低速运行服务率 μ_l	0.2
高速运行服务率 μ_h	0.4
能耗因子 k	10

图 3(a)和 3(b)分别表示在不同激活阈值 θ_3 下, 工作休眠服务率 μ_s 及工作休眠参数 δ 对用户请求平均时延 $E[W]$ 的影响。

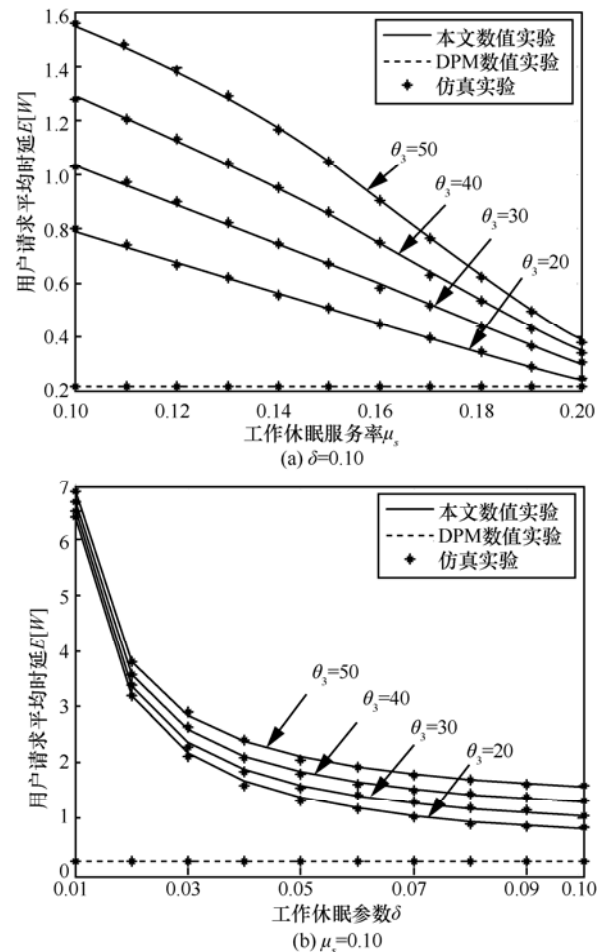


图 3 用户请求平均时延的变化趋势

对图 3(a)进行横向对比, 随着工作休眠服务率 μ_s 的增加, 用户请求平均时延 $E[W]$ 呈下降趋势。工作休眠服务率越大, 虚拟机处于状态 II 时的服务率越大, 即系统整体的服务率越大。因此, 系统缓存中用户请求的数量减少, 用户请求平均时延下降。对图 3(b)进行横向对比, 随着工作休眠参数 δ 的增大, 用户请求平均时延 $E[W]$ 呈下降趋势。工作休眠参数越大, 工作休眠定时器的平均时间长度越短, 备用模块中的虚拟机恢复至高速运行状态的概率越大, 高速运行服务率大于工作休眠服务率。因此, 用户请求平均时延下降。对图 3(a)与图 3(b)进行纵向对比, 固定工作休眠参数 $\delta=0.10$ 及工作休眠服务率 $\mu_s=0.10$, 随着激活阈值 θ_3 的增大, 用户请求平均时延 $E[W]$ 呈上升趋势。激活阈值的增大降低了备用模块中的虚拟机由工作休眠状态恢复至

高速运行状态的概率，故用户请求平均时延增加。

图 4(a)和图 4(b)分别表示在不同激活阈值 θ_3 下，工作休眠服务率 μ_s 及工作休眠参数 δ 对系统节能水平 S 的影响。

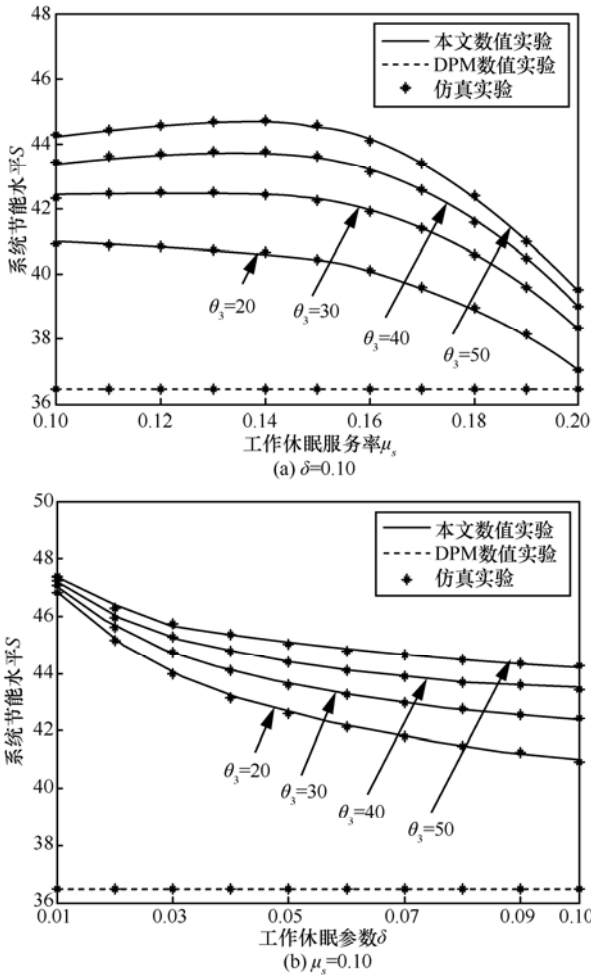


图 4 系统节能水平的变化趋势

对图 4(a)进行横向对比，随着工作休眠服务率 μ_s 的增加，系统节能水平 S 呈先上升后下降趋势，且激活阈值 θ_3 越大，上升的趋势越明显。工作休眠状态服务率越大，在一定的激活阈值下，备用模块中的虚拟机处于工作休眠状态的概率越大，处于高速运行状态的概率越小。单位时间内虚拟机高速运行的能耗水平大于其工作休眠的能耗水平，故备用模块能耗水平下降，系统节能水平增加。当工作休眠服务率增大至一定值后，备用模块中的虚拟机将一直处于工作休眠状态，且随着工作休眠服务率的增加，单位时间内虚拟机的能耗水平也不断增加。因此，备用模块能耗水平增加，系统节能水平下降。图 4(b)中，横向对比，随着工作休眠参数 δ 的增加，

系统节能水平呈下降趋势。工作休眠参数越大，备用模块中的虚拟机工作休眠的平均时间越短，处于高速运行状态的概率越大。高速运行状态的能耗水平大于工作休眠状态的能耗水平，故备用模块能耗水平增加，系统节能水平下降。对图 4(a)与图 4(b)进行纵向对比，固定工作休眠参数 $\delta = 0.10$ 及工作休眠服务率 $\mu_s = 0.10$ ，随着激活阈值 θ_3 的增大，系统节能水平呈上升趋势。激活阈值的增大延长了备用模块中的虚拟机处于工作休眠状态的时间，推迟了其恢复至高速运行状态的时间。因此，备用模块能耗水平下降，系统节能水平增加。

综合上述的实验结果可以看出，与只具有速率调整机制的 DPM 策略相比，本文所提虚拟机调度策略具有更好的节能效果，但系统响应性能有所削弱。在本文的虚拟机调度策略中，合理设置激活阈值，工作休眠服务率及工作休眠参数，可以在保障响应性能的前提下实现能量消耗的有效节约。当激活阈值和工作休眠服务率固定时，需要优化工作休眠参数平衡系统的响应性能和节能效果。

5 工作休眠参数优化

通过权衡不同的系统性能指标，建立系统效用函数^[14] $F(\delta)$ 为

$$F(\delta) = f_s S - f_w E[W] \quad (10)$$

其中， f_s 和 f_w 分别是系统节能水平和用户请求平均时延的影响因子。

沿用第 4 节性能指标的实验参数，并设置 $f_s = 1.0$ 和 $f_w = 1.0$ ，进行系统效用函数的数值实验。实验结果如图 5 所示。

从图 5 可以看出，当工作休眠服务率 μ_s 和激活阈值 θ_3 固定时，随着工作休眠参数 δ 的增大，系统效用函数 $F(\delta)$ 呈现先增大后减小的趋势。当工作休眠参数比较小时，用户请求平均时延的下降趋势比系统节能水平的下降趋势大，即用户请求平均时延占主导因素，故系统效用函数呈现上升趋势。当工作休眠参数比较大时，系统节能水平的下降趋势比用户请求平均时延的下降趋势大，即系统节能水平占主导因素，故系统效用函数呈现下降趋势。因此，当工作休眠参数设置为最优值 δ^* 时，存在最大的系统效用函数 $F(\delta^*)$ 。

由第 3 节的模型解析可知，用户请求平均时延 $E[W]$ 和系统节能水平 S 的封闭解很难给出，系统

效用函数的严格单调性也无法确定。因此，传统的数值算法或分析方法都不适用于虚拟机调度策略中工作休眠参数的优化问题。

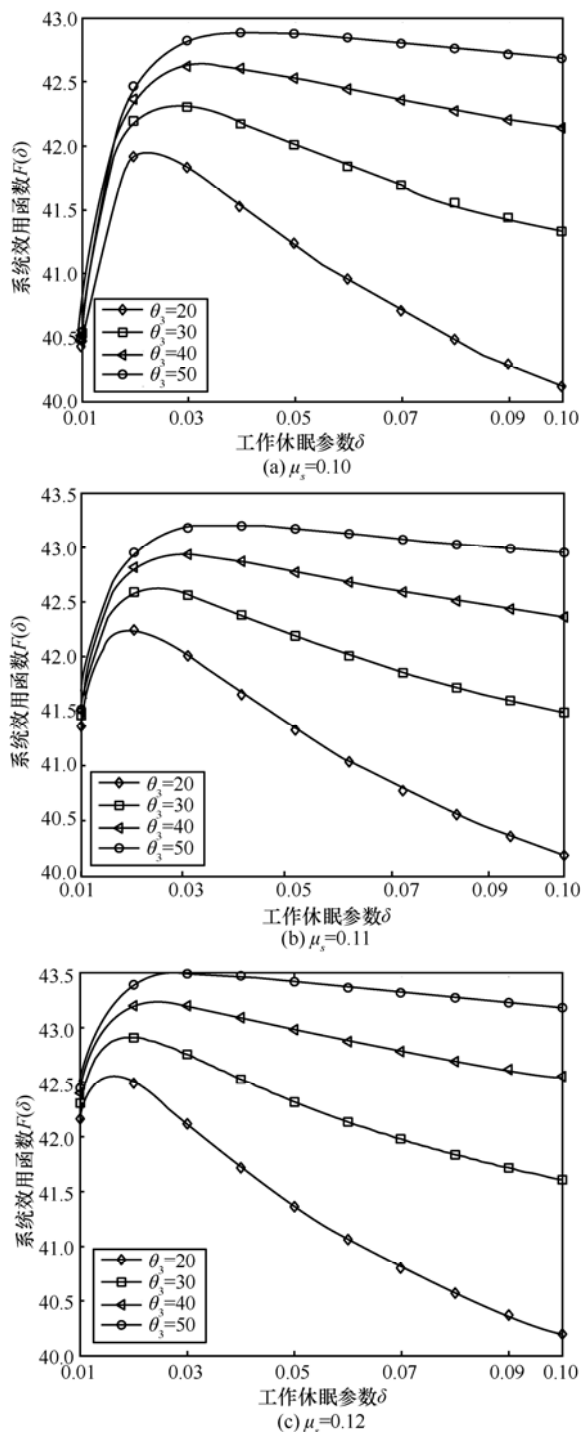


图 5 系统效用函数的变化趋势

为了获得最优工作休眠参数 δ^* 和最大系统效用函数 $F(\delta^*)$ ，借助于仿生群智能优化算法。萤火虫算法^[15,16]是模拟自然界中萤火虫个体之间通过发

光吸引异性的行为构造出的随机优化算法。用搜索空间的点模拟萤火虫个体的位置；将目标函数值模拟成与个体位置对应的自身亮度；将搜索和优化过程模拟成个体吸引和移动的过程，将此过程中用较好可行解取代较差可行解的操作模拟成发光弱的个体向发光强的个体移动。萤火虫算法的求解精度与萤火虫位置的初始值密切相关。混沌系统的遍历性和随机性可以使萤火虫的初始分布更加均匀，应用于智能寻找过程可提高和改善初始群体在搜索空间上的分布质量，加强全局搜索能力。本文利用混沌方程初始化萤火虫的位置，给出一种改进的萤火虫算法，用于求解工作休眠参数 δ 的优化问题。该算法的主要步骤如下所示。

Step1 初始化萤火虫的数量 N ，萤火虫位置的更新次数(内层循环次数) L ，最优解的个数(外层循环次数) K ，混沌因子 ξ ，光强吸收系数 γ ，最大吸引度 β_0 ，步长因子 α 。设当前最优解的序号 $x=1$ 。

Step2 在约束条件 $[0.01,0.10]$ 内，利用混沌方程初始化萤火虫的位置 δ 。 δ_h 表示第 h 只萤火虫的位置， $h \in \{1,2,L,N\}$ 。

$$\delta_1 = 0.09rand + 0.01$$

// rand 表示 $[0,1]$ 间服从均匀分布的随机数

for $h = 2 : N$

$$\delta_h = \xi(\delta_{h-1} - 0.01)(0.10 - \delta_{h-1}) + 0.01$$

end for

Step3 计算第 h 只萤火虫的自身亮度 $I_0(h)$ ， $h \in \{1,2,L,N\}$ 。

$$I_0(h) = F(\delta^h) = f_s S - f_w E[W]$$

// S 和 $E[W]$ 分别是工作休眠参数为 δ_h 时系统节能水平和用户请求平均时延

Step4 令萤火虫位置当前的更新次数 $t=1$ 。计算第 j 只萤火虫相对第 i ， $i, j \in \{1,2,L,N\}$ 只萤火虫的亮度 $I(j,i)$ 。如果此相对亮度大于第 i 只萤火虫的自身亮度，计算第 j 只萤火虫对第 i 只萤火虫的吸引度 $\beta(j,i)$ 和第 i 只萤火虫向第 j 只萤火虫移动的距离 Δ_i ，更新第 i 只萤火虫的位置和自身亮度。

while $t \leq L$

for $i = 1 : N$

for $j = 1 : N$

$$I(j,i) = I_0(j)e^{-\gamma|\delta_j - \delta_i|}$$

if $I(j,i) > I_0(i)$

$$\beta(j,i) = \beta_0 e^{-\gamma|\delta_j - \delta_i|}$$

$$\Delta_i = \beta(j,i)(\delta_j - \delta_i) + \alpha(rand - 0.5) \quad \delta_i = \delta_i + \Delta_i$$

$$I_0(i) = F(\delta_i) = f_s S - f_w E[W]$$

$$t = t + 1$$

end if
end for
end while

Step5 选出目前自身亮度最强的萤火虫

$$\delta[x] = \arg \max_{h \in \{1, 2, L, N\}} \{I_0(h)\}$$

 $x = x + 1$
 // $\delta[x]$ 是一个存储最优解的数组

Step6
 if $x \leq K$
 跳转到 Step 4
 else

$$\delta^* = average(\delta[x]), \quad x \in \{1, 2, L, K\}$$

$$F(\delta^*) = f_s S - f_w E[W]$$

 // S 和 $E[W]$ 分别是工作休眠参数为 δ^* 时系统节能水平和用户请求平均时延

end if

Step7 输出 δ^* 和 $F(\delta^*)$
 设置 $N = 10^4, L = 10^2, K = 100, \xi = 10, \gamma = 1, \beta_0 = 0.01, \alpha = 0.20$, 利用给出的优化算法, 获得最优工作休眠参数 δ^* 和最大系统效用函数 $F(\delta^*)$ 如表 3 所示。

表 3 工作休眠参数 δ^* 的优化结果

μ_s	θ_3	δ^*	$F(\delta^*)$
0.10	20	0.022 7	41.988 3
	30	0.027 2	42.356 8
	40	0.033 3	42.671 5
	50	0.042 2	42.931 8
0.11	20	0.019 5	42.248 3
	30	0.023 4	42.616 4
	40	0.028 9	42.934 2
	50	0.037 0	43.198 8
0.12	20	0.016 1	42.547 4
	30	0.019 4	42.908 2
	40	0.024 1	43.224 6
	50	0.031 2	43.491 4

6 结束语

针对云数据中心产生的空闲能耗和奢侈能耗, 从提高能源效率和实现绿色云计算的角度出发, 提出了一种融合双速率和工作休眠的虚拟机调度策略。通过建立具有自适应服务率和多重工作休假的二维连续时间马尔可夫随机模型, 给出用户请求平均时延、系统节能水平等性能指标。系统理论及仿真实验结果验证了所提策略在保证系统响应性能的前提下可以有效降低系统能耗水平。从经济学的角度构造了系统效用函数, 给出了一种改进的萤火虫算法, 解决了策略中工作休眠参数的优化问题, 实现了系统不同指标之间的平衡。

本文为建立绿色云数据中心提供了一种新思路, 为云数据中心的性能优化提供了理论依据。下一步工作中, 将考虑虚拟机的启动阶段及监听过程, 扩展虚拟机可能的状态, 进一步研究虚拟机调度策略及性能优化问题。

参考文献:

- [1] JIN X, ZHANG F, VASILAKOS A V, et al. Green data centers: a survey, perspectives, and future directions[J]. arXiv: 1608.00687.
- [2] DUAN L, ZHAN D, HOHNERLEIN J. Optimizing cloud data center energy efficiency via dynamic prediction of CPU idle intervals[C]// IEEE International Conference on Cloud Computing. 2015: 985-988.
- [3] LI K. Improving multicore server performance and reducing energy consumption by workload dependent dynamic power management[J]. IEEE Transactions on Cloud Computing, 2016, 4(2): 122-137.
- [4] WANG Y, XIE Q, AMMARI A, et al. Deriving a near-optimal power management policy using model-free reinforcement learning and Bayesian classification[C]// IEEE Design Automation Conference. 2011: 41-46.
- [5] GAO Y, GUAN H, QI Z, et al. Quality of service aware power management for virtualized data centers[J]. Journal of Systems Architecture, 2013, 59(5): 245-259.
- [6] LIAO D, LI K, SUN G, et al. Energy and performance management in large data centers: a queuing theory perspective[C]// International Conference on Computing, Networking and Communications. 2015: 287-291.
- [7] SHEN D, LUO J, DONG F, et al. Stochastic modeling of dynamic right-sizing for energy-efficiency in cloud data centers[J]. Future Generation Computer System, 2015, 48(C): 82-95.
- [8] 马晓彤, 金顺福, 刘建平, 等. 认识无线网络中的基站节能策略及纳什均衡研究[J]. 通信学报, 2017, 37(7): 172-181.
- MA X T, JIN S F, LIU J P, et al. Study on energy saving strategy and Nash equilibrium of base station in cognitive radio network[J]. Journal

on Communications, 2017, 37(7): 172-181.

[9] CHOU C H, WONG D, BHUYAN L N. DynSleep: fine-grained power management for a latency-critical data center application[C]//The 2016 International Symposium on Low Power Electronics and Design. 2016: 212-217.

[10] CHEN Y L, CHANG M F, LIANG W Y, et al. Performance and energy efficient dynamic voltage and frequency scaling scheme for multicore embedded system[C]//IEEE International Conference on Consumer Electronics. 2016: 58-59.

[11] TIAN N, ZHANG Z G. Vacation queueing models theory and applications[M]. Springer US, 2006.

[12] LATOUCHE G, RAMASWAMI V. Introduction to matrix analytic methods in stochastic modeling[M]//Society for Industrial and Applied Mathematics. 2001.

[13] GREENBAUM A. Iterative methods for solving linear systems[M]// Society for Industrial and Applied Mathematics. 2001.

[14] 陈赓, 夏玮玮, 沈连丰. 基于传输速率自适应的动态带宽分配算法[J]. 通信学报, 2014, 35(5): 25-32.
CHEN G, XIA W W, SHEN L F. Dynamic bandwidth allocation algorithm based on transmission rate adaptation[J]. Journal on Communications, 2014, 35(5): 25-32.

[15] GANDOMI A H, YANG X S, ALAVI A H. Mixed variable structural optimization using firefly algorithm[J]. Computers & Structures, 2011, 89(24): 2325-2336.

[16] YU S, ZHU S, MA Y, et al. Enhancing firefly algorithm using generalized opposition-based learning[J]. Computing, 2015, 97(7): 741-754.

作者简介:



金顺福 (1966-), 女, 朝鲜族, 内蒙古满洲里人, 燕山大学教授、博士生导师, 主要研究方向为网络资源分配与优化、排队论研究与应用等。



郝闪闪 (1993-), 女, 河北邢台人, 燕山大学硕士生, 主要研究方向为云数据中心的节能策略与性能优化。



王宝帅 (1993-), 男, 河北保定人, 燕山大学硕士生, 主要研究方向为网络资源分配与优化。