

基于休眠模式与信息可视的云用户接入阈值策略的研究^①

安红宇^② * * * * * 金顺福^③ * * * * * 朱 静 * * * * *

(* 燕山大学信息科学与工程学院 秦皇岛 066004)

(** 河北省计算机虚拟技术与系统集成重点实验室 秦皇岛 066004)

(*** 河北省软件工程重点实验室 秦皇岛 066004)

摘 要 休眠模式广泛应用于绿色云计算环境下的虚拟机(VM)集群中。为了揭示云用户的纳什均衡行为,最大化云系统的社会收益,针对周期性休眠机制,面向云用户触发的网络业务,研究接入阈值策略。在完全可视条件下,求解新到达云用户的预期逗留时间,并建立基于线性收支结构的个人收益函数,得出云用户个人最优接入阈值。面向唤醒状态和休眠状态设置不同的缓存容量,构建多重同步休假模型。运用高斯赛德尔方法,给出排队模型的稳态解,得出云用户社会最优接入阈值。数值实验揭示出在不同虚拟机状态下云用户社会最优接入阈值均小于个人最优接入阈值,为优化云计算环境下的用户接入控制提供了理论依据。

关键词 绿色云计算; 休眠模式; 完全可视排队; 收益函数; 接入阈值

0 引 言

近年来,云计算发展迅速,其应用遍布于通讯、教育、政府、医疗、电力等行业。随着云数据中心规模的扩大和云用户数量的增多,云用户对云系统的服务质量提出了更高的要求,同时,能源消耗也成为日益严峻并备受关注的问题^[1]。

针对云系统中的节能问题,国内外学者进行了相关研究,并取得了丰硕成果。为了降低云计算环境下虚拟机(virtual machine, VM)的能耗,Wattanasomboon 等人^[2]提出了能源感知虚拟机布置(energy-aware virtual machine placement, EVP)方法,通过建立能耗模型评价了 EVP 方法的性能。Wang 等人^[3]提出了一种基于最小能耗的虚拟机资源分配算法,通过引入基于分散多代理(multiagent, MA)的虚拟机分配方法,降低了云计算系统中虚拟机的迁

移成本。针对异构用户服务质量(quality of service, QoS)需求不同及云数据中心能源浪费等问题,金顺福等人^[4]提出了融合区分服务和速率调整的请求调度策略,并通过实验验证了策略的有效性。为了解决能源浪费和资源闲置问题,Chen 等人^[5]提出了动态云资源分配算法。基于 OpenStack 建立基础架构平台,使用配电单元(power distribution unit, PDU)监控系统状态并记录功耗,验证了所提算法在能量节省和云资源规划方面的有效性。Jiang 等人^[6]提出了自适应阈值迁移(adaptive thresholds migration, ATM)算法。根据云资源使用情况评估集群负载,确定动态阈值,面向集群中的空闲和过载节点,使用 ATM 算法确定需要被迁移的虚拟机,平衡资源的使用量和传输开销,优化选择目标主机。这些研究专注于云环境下的节能策略,但没有考虑云系统的社会收益。

基于博弈思想的排队经济学理论已开始应用于

① 国家自然科学基金(61872311,61472342)和河北省自然基金(F2017203141)资助项目。
② 女,1995 年生,硕士生;研究方向:计算机通信网络的系统建模与性能分析;联系人,E-mail: ahy1212@163.com
③ 通信作者,E-mail: jsf@ysu.edu.cn
(收稿日期:2019-04-08)

云计算环境中。为满足云供应商收益最大化的需求,同时保证云用户的最佳服务体验,Baranwal 等人^[7]运用博弈论的思想建立了一种云请求接入控制模型,并基于 CloudSim 模拟器进行了云请求接入控制的性能研究。为了提高云请求的成功执行率,并降低云虚拟机的执行成本与节点收益的公平性偏差 (fairness deviation of profits, FDoP), Xu 等人^[8]研究了云计算环境下虚拟机的调度问题。通过建立多目标优化模型,给出了云用户和云供应商的激励机制。为了有效解决多租户对可分割云资源分配的竞争问题,张晓丹等人^[9]提出一种基于效用函数的最优化分配策略。通过建立评估函数和代价函数,构建效用最优化模型,实现了多租户环境下整体社会收益的最大化。Do 等人^[10]研究了云服务提供商 (cloud service providers, CSPs) 在异构云市场中的价格竞争问题。将博弈模型运用到云计算异构 CSPs 市场和服务选择环境中,通过设定其服务价格以最大化收入。采用种群进化算法和强化学习算法使云用户数量达到均衡状态。考虑移动设备的计算资源限制,以亚马逊弹性计算云的价格信息作为测试数据,Zhang 等人^[11]使用随机博弈方法,提高了云代理和本地云服务的收益。上述成果专注于云系统收益最大化的研究,而忽略了云系统的节能问题。

基于此,综合考虑云系统的能耗节省和云系统的社会收益最大化,在周期性休眠机制的基础之上,面向云用户触发的网络业务,研究接入阈值策略。在完全可视情况下,云用户根据当前系统中的云用户数量和云虚拟机状态,决定是否接入云系统。通过求解新到达云用户的预期逗留时间及预期个人收益,得出个人最优接入阈值。建立多重同步休假模型,运用高斯赛德尔法求出系统的稳态解,计算单位时间的社会收益,给出社会最优接入阈值。通过数值实验揭示出社会最优接入阈值小于个人最优接入阈值的规律。

1 休眠机制与系统模型

1.1 周期性休眠机制

云数据中心的虚拟机长期保持开启状态将产生

大量的能源浪费。作为一种虚拟化资源管理技术,休眠机制成功应用于云数据中心的节能策略中,有效提高了云数据中心的能源效率。

为了解决能耗过大和负载不均衡的问题,往往以集群的方式管理云数据中心的虚拟机。相比于单重休眠机制^[12],周期性休眠机制在云数据中心具有更好的节能效果。假设云用户同构,集群内的虚拟机也同构,在一个集群内的虚拟机可以同步工作在周期性休眠机制下。当所有云用户都完成服务时,集群内的全部云虚拟机同时进入休眠状态 ($j = 0$),开始一个休眠周期,该休眠周期的长度由休眠定时器控制。当定时器超时后,若缓存仍为空,集群内所有云虚拟机进入下一个休眠周期;否则同时进入唤醒状态 ($j = 1$)。当系统中的云用户数量 l ($l = 1, 2, \dots$) 小于集群内云虚拟机数量 k ($k = 1, 2, \dots$), 即 $l < k$ 时,随机选择 l 个云虚拟机服务全部云用户,其余 $k - l$ 个云虚拟机则处于空闲状态;否则按照到达顺序服务系统中的前 k 个云用户,另外 $l - k$ 个云用户继续在系统中等待。

1.2 系统模型

本文考虑云用户触发的网络业务^[13],假设云用户的到达服从参数为 λ 的泊松过程。服务一个云用户的时间服从参数为 μ 的指数分布。假设休眠定时器的长度服从参数为 θ 的指数分布,参数 θ 也称为休眠参数。假设云用户的到达间隔、服务时间以及休眠定时器长度是相互独立的。

令 $L(t)$ 表示时刻 t 系统中的云用户数量; $J(t)$ 表示时刻 t 集群内云虚拟机所处的状态, $J(t) = 0$ 表示云虚拟机处于休眠状态, $J(t) = 1$ 表示云虚拟机处于唤醒状态。在上述假设条件下,云系统状态 $(L(t), J(t))$ 构成连续时间 Markov 模型。

在完全可视情况下,每个到达系统的云用户都会得知系统中云用户数量 $L(t)$ 和云虚拟机状态 $J(t)$, 云用户据此信息决定是否接入系统接受服务。假设每个云用户完成服务后获得的收益为 R , 逗留在系统中单位时间的等待费用为 C 。考虑云用户是风险中立的,一旦决定接入系统,不允许中途退出,已决定止步的云用户也不允许重新返回系统。

为了揭示绿色云环境下云用户的纳什均衡行

为,最大化云系统的社会收益需研究可视条件下云用户的接入阈值策略,包括云用户的个人最优接入阈值策略和社会最优接入阈值策略。

2 个人最优接入阈值策略

在周期性同步休眠机制下,考虑系统中云用户数量和集群内云虚拟机数量的关系,研究云用户的个人最优接入阈值策略。

在系统中云用户的数量 l 小于集群内云虚拟机的数量 k 的条件下,分析云用户的接入行为。

当云虚拟机处于休眠状态 ($j = 0$) 时,选择接入云系统的云用户预期逗留时间:

$$S_1(0) = \frac{1}{\theta} + \frac{1}{\mu} \quad (1)$$

选择接入云系统的云用户预期个人收益:

$$B_{i1}(0) = R - C\left(\frac{1}{\theta} + \frac{1}{\mu}\right) \quad (2)$$

当云虚拟机处于唤醒状态 ($j = 1$) 时,选择接入云系统的云用户预期逗留时间:

$$S_1(1) = \frac{1}{\mu} \quad (3)$$

选择接入云系统的云用户预期个人收益:

$$B_{i1}(1) = R - \frac{C}{\mu} \quad (4)$$

为了保证集群中存在空闲云虚拟机的条件下到达的云用户一定选择接入系统,则有:

$$B_{i1}(0) > 0, B_{i1}(1) > 0 \quad (5)$$

即:

$$R > C\left(\frac{1}{\theta} + \frac{1}{\mu}\right) \quad (6)$$

当集群内所有云虚拟机均被占用时,新到达的云用户将根据系统中云用户的数量和云虚拟机的状态计算其预期逗留时间及预期个人收益。只有当个人收益非负时,新到达的云用户才接入系统接受服务,否则,离开系统。

在系统中云用户的数量 l 不小于集群内云虚拟机的数量 k 的条件下,若云虚拟机处于休眠状态 ($j = 0$),新到达的云用户一旦选择接入系统,其预期逗留时间为

$$S_2(0) = \frac{1}{\theta} + \frac{1}{k\mu}(l - k + 1) + \frac{1}{\mu} \quad (7)$$

预期个人收益:

$$\begin{aligned} B_{i2}(0) &= R - CS_2(0) \\ &= R - C\left(\frac{1}{\theta} + \frac{1}{k\mu}(l - k + 1) + \frac{1}{\mu}\right) \end{aligned} \quad (8)$$

令 $B_{i2}(0) = 0$, 可得系统状态处于 $(l, 0)$ 下的个人最优接入阈值策略:

$$n_e(0) = \lfloor \frac{Rk\mu}{C} - \frac{k\mu}{\theta} \rfloor - 1 \quad (9)$$

若云虚拟机处于唤醒状态 ($j = 1$), 新到达的云用户一旦选择接入系统,其预期逗留时间为

$$S_2(1) = \frac{1}{k\mu}(l - k + 1) + \frac{1}{\mu} \quad (10)$$

预期个人收益为

$$\begin{aligned} B_{i2}(1) &= R - CS_2(1) \\ &= R - C\left(\frac{1}{k\mu}(l - k + 1) + \frac{1}{\mu}\right) \end{aligned} \quad (11)$$

令 $B_{i2}(1) = 0$, 可得系统状态处于 $(l, 1)$ 下的个人最优接入阈值策略:

$$n_e(1) = \lfloor \frac{Rk\mu}{C} \rfloor - 1 \quad (12)$$

综合上述策略可得,在云虚拟机处于休眠状态的条件下,如果新到达云用户得知系统中当前云用户数量 $l \leq n_e(0)$, 则选择接入系统,否则,选择离开。

在云虚拟机处于唤醒状态的条件下,如果新到达云用户得知系统中当前云用户数量 $l \leq n_e(1)$, 则选择接入系统,否则,选择离开。

3 社会最优接入阈值策略

在个人最优接入阈值策略下,每个接入系统的云用户均可获得收益。为了保证系统整体的收益最优,需进一步研究社会最优接入阈值策略。

当集群内的云虚拟机处于休眠状态时,若到达系统的云用户得知当前系统中的云用户数量为 $n_e(0) + 1$, 则选择离开系统,即系统中最多有 $n_e(0) + 1$ 个云用户。当集群中的云虚拟机处于唤醒状态时,若到达系统的云用户得知当前系统中的云用户

数量为 $n_e(1) + 1$, 则选择离开系统, 即系统中最多有 $n_e(1) + 1$ 个云用户。

据此, Markov 链 $\{(L(t), J(t)): t \geq 0\}$ 的状态空间为 $\Omega = \{(l, 0) | 0 \leq l \leq n_e(0) + 1\} \cup \{(l, 1) | 0 \leq l \leq n_e(1) + 1\}$ 。令 $\pi_{l,j}$ 表示该 Markov 链的稳态分布, 即:

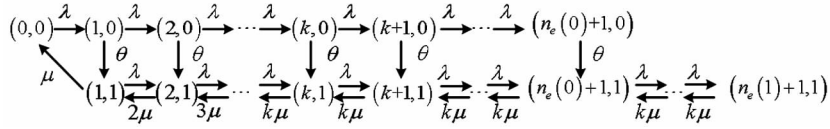


图 1 系统的状态转移过程

根据图 1 给出的系统状态转移过程, 建立如下方程组:

$$\begin{cases} \lambda \pi_{0,0} = \mu \pi_{1,1} \\ \lambda \pi_{l-1,0} = (\lambda + \theta) \pi_{l,0}, & 1 \leq l \leq n_e(0) \\ \lambda \pi_{n_e(0),0} = \theta \pi_{n_e(0)+1,0} \\ \lambda \pi_{l-1,1} + \theta \pi_{l,0} - (\lambda + l\mu) \pi_{l,1} + (l+1)\mu \pi_{l+1,1} = 0, & 2 \leq l < k \\ \lambda \pi_{l-1,1} + \theta \pi_{l,0} - (\lambda + k\mu) \pi_{l,1} + k\mu \pi_{l+1,1} = 0, & k \leq l \leq n_e(0) + 1 \\ \lambda \pi_{l-1,1} - (\lambda + k\mu) \pi_{l,1} + k\mu \pi_{l+1,1} = 0, & n_e(0) + 1 < l \leq n_e(1) \\ \lambda \pi_{n_e(1),1} - k\mu \pi_{n_e(1)+1,1} = 0 \end{cases} \quad (14)$$

采用高斯赛德尔方法^[14], 结合正规化条件, 可得 Markov 链 $\{(L(t), J(t)): t \geq 0\}$ 的稳态解 $\pi_{l,j}$ 。

从社会的角度考虑整个系统, 定义社会收益为系统中所有云用户和云服务商的总收益。云用户与云服务商之间的支付转移对于整个系统的优化没有影响, 因此, 单位时间的社会收益 B_s 可表述为系统中所有云用户完成服务后获得的回报减去总的等待费用^[15]。

在个人最优接入阈值策略 $(n_e(0), n_e(1))$ 的限制下, 单位时间的社会收益为

$$B_s = R\lambda(1 - \pi_{n_e(0)+1,0} - \pi_{n_e(1)+1,1}) - C\left(\sum_{l=0}^{n_e(0)+1} l\pi_{l,0} + \sum_{l=0}^{n_e(1)+1} l\pi_{l,1}\right) \quad (15)$$

最大社会收益 B_s 所对应的接入阈值称为社会

$$\pi_{l,j} = \lim_{t \rightarrow \infty} P\{L(t) = l, J(t) = j\}, (l, j) \in \Omega \quad (13)$$

基于 1.1 节系统模型的工作机制, 得到系统状态转移图, 如图 1 所示。

最优接入阈值, 即:

$$(n^*(0), n^*(1)) = \operatorname{argmax}\{B_s\} \quad (16)$$

4 系统实验

由于云用户的社会最优接入阈值解析式无法用显示形式给出, 通过系统实验研究完全可视情形下社会收益的变化趋势, 并揭示个人最优接入阈值和社会最优接入阈值之间的偏差。

系统实验在交互式软件开发环境 Matlab R2016a^[16]下完成。

在约束条件 $R > C\left(\frac{1}{\mu} + \frac{1}{\theta}\right)$ 下, 设置实验参数如表 1 所示。

表 1 实验参数设置

系统参数	参数值
集群中云虚拟机总数 k	5 台
云用户到达率 λ	3.10 个/ms
云虚拟机服务率 μ	1.35 个/ms
云虚拟机唤醒率 θ	0.08 次/ms
云用户的等待费用 C	10
云用户完成服务后的收益 R	50

为了直观地揭示云用户在休眠状态下的接入阈值 $n(0)$ 和唤醒状态下的接入阈值 $n(1)$ 对社会收益的影响, 用 3 维图形给出单位时间社会收益 B_s 的变化趋势, 如图 2 所示。

由图 2 可知, 当休眠状态下的云用户接入阈值

$n(0)$ 一定时,随着唤醒状态下云用户接入阈值 $n(1)$ 的增大,单位时间的社会收益 B_s 显现出先增大再减小的变化规律。在云虚拟机服务能力的范围内,随着接入阈值的增大,接入云系统的云用户增多,单位时间的社会收益增大。当接入阈值增大到一定程度时,系统中的云用户数量过多,造成云用户的平均等待费用增大,所以单位时间的社会收益将逐渐减小。同理,当唤醒状态下云用户接入阈值 $n(1)$ 一定时,随着休眠状态下云用户接入阈值 $n(0)$ 的增大单位时间的社会收益也显现出先增大后减小的变化规律。

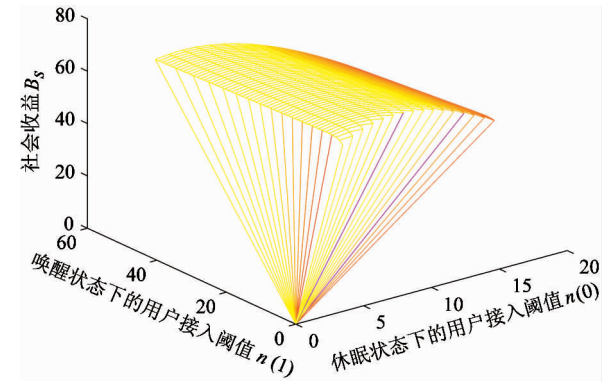


图 2 单位时间社会收益的 3 维变化趋势

为了找到单位时间社会收益的最大值及最大值所对应的云用户接入阈值,以等高线的形式给出单位时间社会收益 B_s 的变化趋势,如图 3 所示。

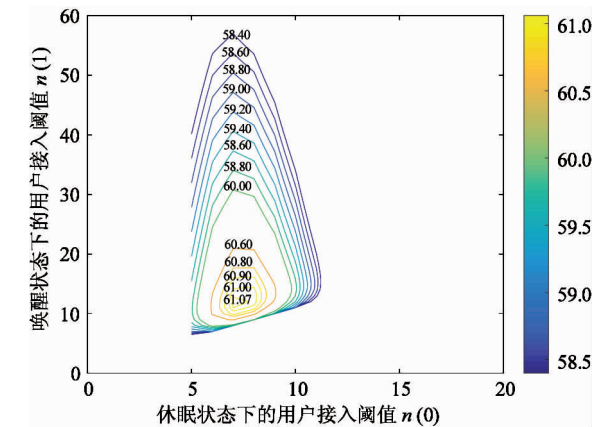


图 3 单位时间社会收益的等高线变化趋势

由图 3 可知,当单位时间的社会收益取得最大值 $B_s^* = 61.07$ 时,云用户的社会最优接入阈值为 $(n^*(0), n^*(1)) = (7, 11)$ 。该结果表明,云用户

在唤醒状态下的社会最优接入阈值大于休眠状态下的社会最优接入阈值。

沿用相同的实验参数,利用第 2 节中的式(9)和式(12),求得云用户的个人最优接入阈值 $(n_e(0), n_e(1)) = (33664, 33749)$ 。同样可以看出,云用户在唤醒状态下的个人最优接入阈值大于休眠状态下的个人最优接入阈值。

对比个人最优接入阈值与社会最优接入阈值,可知,在不同的云虚拟机状态下,云用户的社会最优接入阈值远小于个人最优接入阈值,即 $n_e(0) > n^*(0)$, $n_e(1) > n^*(1)$ 。基于该规律,云供应商有充分的空间制定状态依赖的收费策略,以实现社会收益的最大化。

5 结 论

在采用周期性休眠机制的云资源管理框架下,面向云用户触发的网络业务,研究了接入阈值策略。考虑云虚拟机状态与系统中云用户数量均可见的情况,面向云虚拟机的唤醒状态和休眠状态,给出了个人最优与社会最优接入阈值策略。实验结果表明,云用户在唤醒状态下的最优接入阈值大于休眠状态下的最优接入阈值,且确定状态下的社会最优接入阈值远小于个人最优接入阈值。该规律为制定云用户的系统接入费用提供了理论依据。在下一步研究中,考虑自相似的网络业务,引入 Markov 到达过程和一般服务过程,建立具有普适意义的系统模型,结合云系统的实测数据,进行系统实验,进一步研究绿色云计算环境下资源调度问题与系统优化的问题。

参考文献

[1] Yang C, Yan Y, Chen S, et al. iGEMS: a cloud green energy management system in data center[C] // Proceedings of the 11th International Conference on Green, Pervasive, and Cloud Computing, Xi'an, China, 2016: 82-98

[2] Wattanasomboon P, Somchit Y. Virtual machine placement method for energy saving in cloud computing[C] // International Conference on Information Technology & Electrical Engineering, Guangzhou, China, 2015: 275-280

[3] Wang W, Jiang Y, Wu W. Multiagent-based resource al-

- location for energy minimization in cloud computing systems[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017, 47(2): 205-220
- [4] 金顺福, 薛元铮, 顾蕊. 融合区分服务和速率调整的请求调度策略[J]. 高技术通讯, 2017, 27(9-10): 790-799
- [5] Chen C C, Sun P L, Yang C T, et al. Implementation of a cloud energy saving system with virtual machine dynamic resource allocation method based on openstack[C] // 7th International Symposium on Parallel Architectures, Nanjing, China, 2016: 190-196
- [6] Jiang C, Wu J, Li Z. Adaptive thresholds determination for saving cloud energy using three-way decisions[J]. *Cluster Computing*, 2018, 1879(7): 1-8
- [7] Baranwal G, Vidyarthi D. Admission control in cloud computing using game theory[J]. *The Journal of Supercomputing*, 2016, 72(1): 317-346
- [8] Xu H, Liu Y, Wei W, et al. Incentive-aware virtual machine scheduling in cloud computing[J]. *The Journal of Supercomputing*, 2018, 74(7): 3016-3038
- [9] 张晓丹, 郑迎凤, 王岁花. 可分割云资源下多租户效用最优策略[J]. 计算机工程与设计, 2017, 38(7): 1854-1858
- [10] Do C T, Tran N H, Huh E N, et al. Dynamics of service selection and provider pricing game in heterogeneous cloud market[J]. *Journal of Network and Computer Applications*, 2016, 69: 152-169
- [11] Zhang F, Liu Y, Liu W, et al. Stochastic game between cloud broker and cloudlet for mobile cloud computing[C] // 2017 IEEE 86th Vehicular Technology Conference (VTC-Fall), Toronto, Canada, 2018: 1-5
- [12] Liu W, Ma Y, Li J. Equilibrium threshold strategies in observable queueing systems under single vacation policy[J]. *Applied Mathematical Modelling*, 2012, 36(12): 6186-6202
- [13] Paxson V, Floyd S. Wide area traffic: the failure of Poisson modeling[J]. *IEEE/ACM Transactions on Networking*, 1995, 3(3): 226-224
- [14] Tian Z, Tian M, Liu Z, et al. The Jacobi and Gauss-Seidel-type iteration methods for the matrix equation[J]. *Applied Mathematics and Computation*, 2017, 292: 63-75
- [15] Hassin R, Haviv M. To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems[M]. Boston/Dordrecht/London: Kluwer Academics Publishers, 2003: 21-33
- [16] Jin S, Wang X, Yue W. A task scheduling strategy with a sleep-delay timer and a waking-up threshold in cloud computing[C] // International Conference on Queueing Theory and Network Applications, Tsukuba, Japan, 2018: 115-123

Research on access threshold strategy of cloud users based on sleep mode and observable information

An Hongyu^{*****}, Jin Shunfu^{*****}, Zhu Jing^{*****}

(* School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004)

(** Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004)

(*** Key Laboratory for Software Engineering of Hebei Province, Qinhuangdao 066004)

Abstract

Sleep mode has been widely applied in virtual machine (VM) clusters in green cloud environments. In order to reveal the Nash equilibrium behavior of cloud users and maximize the social benefit of the cloud system, the access threshold strategy is investigated for the network services triggered by cloud users with a periodic sleep mode. From a fully observable scenario, the expected delay of a newly arriving cloud user is calculated. In addition, a reward-cost structure based benefit function of an individual cloud user is established, as well as the individually optimal access threshold of cloud users is obtained. With respect to the awake state and the sleep state, respectively, a synchronous multiple-vacation queueing model with diverse cache capacities is constructed. By using the Gauss-Seidel method, we derive the steady-state solution of the queueing model and obtain the socially optimal access threshold of cloud users. The experiment results show that the socially optimal access threshold is always smaller than the individually optimal access threshold no matter whether the VMs are asleep or awake. The proposed method provides a theoretical basis for optimizing the access control of users over cloud environment.

Key words: green cloud environment, sleep mode, fully observable queue, benefit function, access threshold