

# 基于休眠模式与注册服务的云架构性能优化问题的研究

金顺福<sup>1,2</sup>, 武海星<sup>1,2</sup>, 霍甜甜<sup>1,2</sup>, 赵文娟<sup>1,2</sup>

(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)

**摘要:** 兼顾能量节省与匿名用户的服务质量, 结合免费云服务与注册云服务, 提出了一种基于休眠模式的新型云架构。将免费云服务抽象为第一次服务, 注册云服务抽象为第二次服务, 休眠抽象为休假, 建立了一个带有二次可选服务且部分虚拟机异步多重休假的排队模型。运用矩阵几何解方法求解排队模型的稳态分布, 研究了系统节能率与匿名用户平均响应时间等性能指标。综合匿名用户的服务收益与等待服务所消耗的时间成本, 建立了收益函数。利用数值结果揭示了匿名用户纳什均衡到达率与社会最优到达率之间的关系。为实现新型云架构下的社会最优提供理论依据。

**关键词:** 云架构; 注册云服务; 休眠模式; 二次可选服务; 纳什均衡; 社会最优

**中图分类号:** TP393

**文献标识码:** A

**doi:** 10.11959/j.issn.1000-436x.2019202

## On the performance optimization for the cloud architecture with sleep-mode and registration service

JIN Shunfu<sup>1,2</sup>, WU Haixing<sup>1,2</sup>, HUO Tiantian<sup>1,2</sup>, ZHAO Wenjuan<sup>1,2</sup>

1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China

2. The Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004, China

**Abstract:** In order to balance the energy conservation and the service quality for anonymous users, a novel sleep-mode based architecture with the free cloud service and the registration cloud service was proposed. Regarding the free cloud service as the first service, the registration cloud service as the second service and the sleep state as the vacation, a partial asynchronous multiple vacation queueing model with a second optional service was built. Applying the method of a matrix-geometric solution, the steady-state distribution of the queueing model was derived, and then the energy saving rate of system as well as the average response time of anonymous users were estimated. By considering the benefits from accessing the cloud service and the time cost on waiting for the cloud service, benefit functions were constructed. Numerical results were carried out to reveal the relationship between Nash equilibrium arrival rate and social optimal arrival rate of anonymous users. The proposed cloud architecture provides a theoretical basis for social optimization.

**Key words:** cloud architecture, registration cloud service, sleep-mode, second optional service, Nash equilibrium, social optimization

### 1 引言

云计算通过互联网向云用户提供服务<sup>[1-2]</sup>。随着云计算需求的不断增长, 云服务的种类也在不断

增加, 云供应商之间的竞争越来越激烈<sup>[3-4]</sup>。为了建立忠实的用户群体, 部分云供应商在提供免费云服务的基础上, 引入了会员制度。另外, 云系统的能量消耗问题日益严重, 绿色云计算成为云系统可持

收稿日期: 2019-01-18; 修回日期: 2019-08-25

基金项目: 国家自然科学基金资助项目 (No.61872311, No.61973261); 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.61872311, No.61973261), The Natural Science Foundation of Hebei Province (No.F2017203141)

续发展的必要条件。因此,如何吸引更多的云用户并有效降低云系统中的能量消耗成为相关领域的研究热点。

会员制度的营销策略广泛应用于多个方面。文献[5]提出了一种基于 RFID (radio frequency identification) 技术的 VIP (very important person) 客户信息服务管理系统,应用于服装零售商店以保证 VIP 用户能够获得更好的服务。文献[6]介绍了一种可以在家庭媒体服务器或个人电脑上轻松实现虚拟私人云服务的管理方案,研究了更具扩展性与实用性的会员管理方法。在开放的云平台上,许多云供应商以提供免费云服务的方式吸引更多的云用户。对免费云服务满意的匿名用户很可能选择注册成为 VIP 用户,以获得更高质量的服务。

云计算系统因其巨大的能量消耗引起人们的普遍关注。文献[7]考虑在空闲计算服务器和网络硬件中启用休眠模式以减少能耗。文献[8]将 DPM (dynamic power management) 技术与同步多重工作休眠模式相结合,提出了一种虚拟机调度策略,降低了云数据中心的能量消耗。以上文献专注于云计算常规服务的节能策略。如何建立一个带有注册服务且考虑能量节省的云架构对云供应商来说非常重要。

随机模型理论是一种有效的系统性能评估手段。带有二次可选服务的排队模型适用于云系统中的会员注册服务。2000年, Madan<sup>[9]</sup>首次提出带有二次可选服务的 M/G/1 的排队系统,使用补充变量法,分析了顾客排队的平均队长及平均等待时间等性能指标。此后,大量的文献研究了带有二次可选服务的排队模型。文献[10]提出了一种带有二次可选服务和随机反馈的二相串联排队模型,通过极大似然估计和贝叶斯估计方法,确定了稳态下系统的到达率、服务率及流通强度等系统参数。文献[11]采用矩阵几何解方法,研究了一个具有二次可选服务和工作故障的 M/M/1 排队系统,利用遗传算法解决了成本优化问题。目前,二次可选服务排队模型的研究局限于单服务台。在云计算等应用的驱动下,多服务台排队模型的研究成为必然趋势。

带有休假机制的排队模型适合于云系统中的休眠模式。休假模型由 Doshi<sup>[12]</sup>首次提出,此后越来越多的学者研究了休假排队模型。文献[13]面向认知无线网络提出了一种基站半休眠模式,通过建

立具有抢占式优先级服务和单重工作休假的排队模型,揭示了系统节能与用户时延间的折中关系。文献[14]考虑了带有服务台故障的异步休假策略,可以帮助系统工程师和决策者提高机械加工系统的可用性。区别于以上研究,文献[15]提出了带有顾客止步和伯努利休假的二次可选服务的 Geo/G/1 排队系统,采用补充变量法,得到了重试组队长和系统队长的概率母函数。考虑带有注册服务与能量节省的云架构,需要研究一种具有二次可选服务和休假机制的多服务台排队模型。

本文提出了一种基于能量节省的带有免费云服务和注册云服务的云架构,免费云服务与注册云服务由同一台虚拟机提供。将云用户分为两类:匿名用户和 VIP 用户。为了获得更高质量的服务,部分匿名用户结束免费云服务后很可能选择注册云服务成为 VIP 用户。从能量节省的角度出发,本文允许部分虚拟机在系统缓存中没有排队等待的匿名用户时进入休眠状态,建立了一个带有二次可选服务且部分虚拟机异步多重休假的排队模型,从系统的节能率和匿名用户的平均响应时间 2 个方面评估系统性能。如果有较多的匿名用户进入系统,匿名用户的响应时间较长,服务质量较低;如果有较少的匿名用户进入系统,系统的节能率较高,但是云供应商的利润较低。因此,需要平衡匿名用户的服务质量与云供应商的利益,研究匿名用户的纳什均衡行为和社会最优行为。云供应商可以通过投放广告等方式使匿名用户的纳什均衡到达率与社会最优到达率一致,最终实现云系统的社会最优。

## 2 新型云架构及系统模型

### 2.1 新型云架构

本文兼顾系统的节能效果及匿名用户的服务性能,将休眠模式引入部分虚拟机中,提出了一种新型云架构,如图 1 所示。

云环境中物理机的处理能力一般很强,在一台物理机上部署多台虚拟机可以提高吞吐量与并行度。为吸引更多用户,云供应商通常向匿名用户提供免费云服务。对免费云服务满意的匿名用户可能选择注册云服务成为 VIP 用户。为了向匿名用户提供及时的注册过程,所提云架构将免费云服务与注册云服务部署在同一台虚拟机上。本文兼顾系统的节能效果及匿名用户的服务性能,引入了阈值  $d$  控制虚拟机状态,为了保证匿名用户的服务性能,所

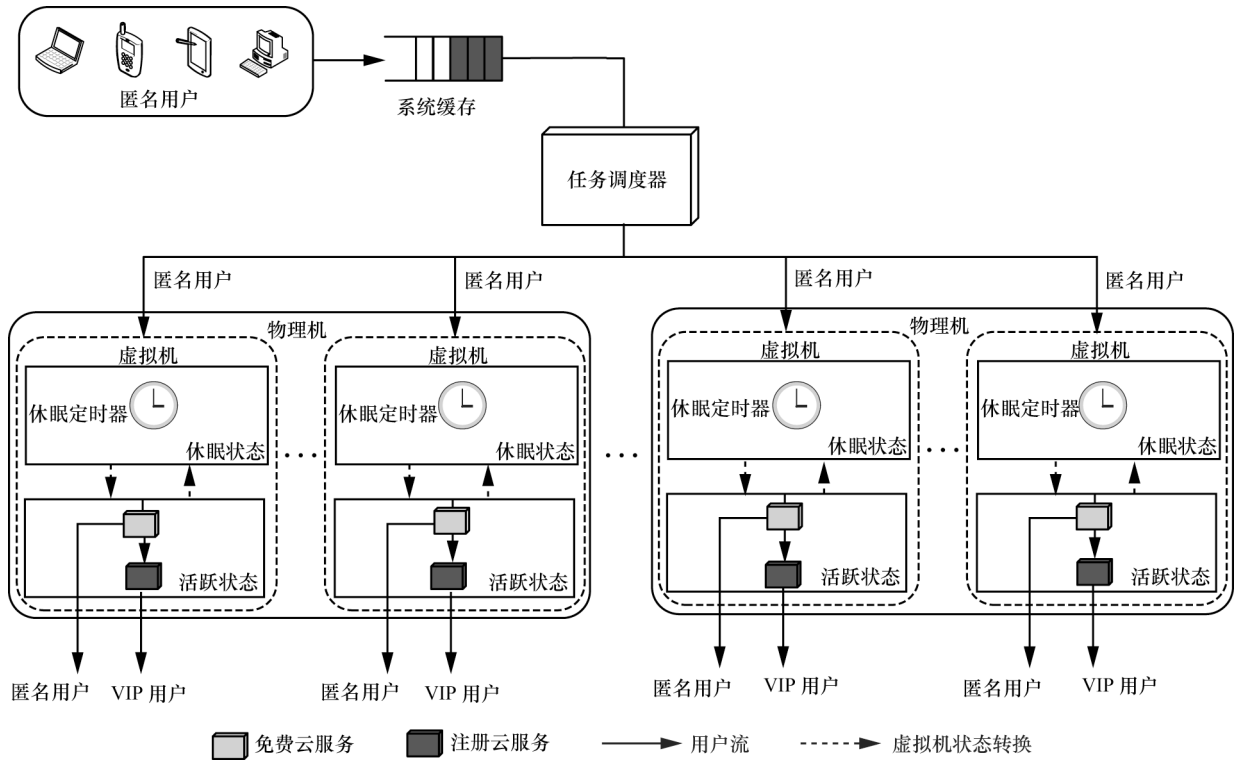


图 1 本文所提云架构示意

提云架构将处于活跃状态的虚拟机数量控制在阈值  $d$  之上。

该云架构中虚拟机状态转移过程如下。

1) 匿名用户进入系统后在系统缓冲区排队等待接受免费服务。一旦处于活跃状态的虚拟机空闲，任务调度器立即将该虚拟机分配给系统缓存中排在队首的匿名用户，分配到虚拟机的匿名用户开始接受免费服务。

2) 综合考虑接受注册云服务带来的好处与花费的时间，完成免费服务的匿名用户可能选择注册云服务，也可能直接离开系统。选择注册云服务的匿名用户在经过注册过程之后转化为 VIP 用户。直接离开系统，放弃注册云服务的匿名用户保持原身份不变。

3) 在一台虚拟机刚刚腾空时，如果系统缓存不空，该虚拟机保持在活跃状态并立即为系统缓存中排在队首的匿名用户提供服务。如果系统缓存为空且系统中处于活跃状态的虚拟机数量已降至阈值  $d$ ，该虚拟机将持续在活跃状态，为即将到达的匿名用户提供服务。如果系统缓存为空且处于活跃状态的虚拟机数量高于阈值  $d$ ，该虚拟机将进入周期性休眠状态。每个休眠周期的长度是随机变化的，由休眠定时器决定。一个休眠周期结束后，如果系

统缓存仍然为空，该虚拟机将继续保持休眠，并开始一个新的休眠周期；否则，该虚拟机将进入活跃状态，为缓存中的匿名用户提供服务。

为了在所建立的云架构下评估云系统性能，需建立系统模型刻画匿名用户的随机行为。

## 2.2 系统模型

基于所提云架构，本节建立带有二次可选服务且部分虚拟机异步多重休眠的排队模型。

参考云系统性能优化相关文献[16,17]，做出以下假设。匿名用户的到达服从参数为  $\lambda(\lambda > 0)$  的泊松过程，休眠定时器的长度服从参数为  $\theta(\theta > 0)$  的指数分布；一个匿名用户的免费服务时间与注册服务时间分别服从参数为  $\mu_1$  和  $\mu_2(\mu_1, \mu_2 > 0)$  的指数分布；一个匿名用户完成免费云服务后选择注册云服务的概率为  $q$ ，直接离开系统的概率为  $\bar{q}(\bar{q} = 1 - q)$ 。则系统的负载强度为

$$\rho = \frac{\lambda(\mu_2 + q\mu_1)}{c\mu_1\mu_2} \quad (1)$$

其中， $c$  为系统中虚拟机总数。系统处于稳定状态的充分必要条件是  $\rho < 1$ 。

令随机变量  $N(t) = i(i = 0, 1, 2, \dots)$  表示  $t$  时刻系统中匿名用户的数量，称为系统水平。令随机变量



2)  $A_x$  表示系统水平  $x$  经过一步转移后保持不变的子阵。

当  $x=0$  时, 系统为空,  $A_0$  退化为一个数值。

$$A_0 = -\lambda \tag{7}$$

当  $1 \leq x \leq d+1$  时, 全部的匿名用户都在接受虚拟机的服务, 即系统中匿名用户的数量与为匿名用户提供服务 (包括免费服务和注册服务) 的虚拟机数量相等。  $A_x$  为  $(x+1) \times (x+1)$  维方阵。

为了进一步分析  $A_x$  的形式, 引入符号  $\alpha_0 = -(\lambda + x\mu_1)$ ,  $\alpha_1 = -(\lambda + (x-1)\mu_1 + \mu_2)$ ,  $\dots$ ,  $\alpha_{x-1} = -(\lambda + \mu_1 + (x-1)\mu_2)$ ,  $\alpha_x = -(\lambda + x\mu_2)$ 。  $A_x$  形式如式(8)所示。

$$A_x = \begin{pmatrix} \alpha_0 & xq\mu_1 & & & & \\ & \alpha_1 & (x-1)q\mu_1 & & & \\ & & \alpha_2 & \ddots & & \\ & & & \ddots & q\mu_1 & \\ & & & & & \alpha_x \end{pmatrix} \tag{8}$$

当  $d+2 \leq x \leq c$  时, 处于活跃状态的虚拟机的数量至少为  $d+1$ , 至多为系统中匿名用户的数量  $x$ 。  $A_x$  为  $\frac{1}{2}(x+d+3)(x-d) \times \frac{1}{2}(x+d+3)(x-d)$  维方阵, 形式如式(9)所示。

$$A_x = \begin{pmatrix} a_{11} & a_{12} & & & & \\ & a_{22} & a_{23} & & & \\ & & \ddots & \ddots & & \\ & & & & & \\ & & & & a_{(x-d-1)(x-d-1)} & a_{(x-d-1)(x-d)} \\ & & & & & a_{(x-d)(x-d)} \end{pmatrix} \tag{9}$$

其中, 元素  $a_{nn}$  ( $1 \leq n \leq x-d$ ) 表示为匿名用户提供服务的虚拟机数量保持在  $d+n$  的转移率子阵。元素  $a_{n(n+1)}$  ( $1 \leq n \leq x-d-1$ ) 表示为匿名用户提供服务的虚拟机数量由  $d+n$  增加到  $d+n+1$  的转移率子阵。  $a_{nn}$  与  $a_{n(n+1)}$  均为  $(d+n+1) \times (d+n+1)$  维方阵。

为了进一步分析  $A_x$  的形式, 引入符号  $\gamma_0 = -(\lambda + (d+n)\mu_1 + (c-d-n)\theta)$ ,  $\gamma_1 = -(\lambda + (d+n-1)\mu_1 + \mu_2 + (c-d-n)\theta)$ ,  $\dots$ ,  $\gamma_{d+n-1} = -(\lambda + \mu_1 + (d+n-1)\mu_2 + (c-d-n)\theta)$ ,  $\gamma_{d+n} = -(\lambda + (d+n)\mu_2 + (c-d-n)\theta)$ ;  $\eta_0 = (d+n)q\mu_1$ ,  $\eta_1 = (d+n-1)q\mu_1$ ,  $\dots$ ,  $\eta_{d+n-2} = 2q\mu_1$ ,  $\eta_{d+n-1} = q\mu_1$ 。  $a_{nn}$  与  $a_{n(n+1)}$  分别为

$$a_{nn} = \begin{pmatrix} \gamma_0 & \eta_0 & & & & \\ & \gamma_1 & \eta_1 & & & \\ & & \ddots & \ddots & & \\ & & & \gamma_{d+n-1} & \eta_{d+n-1} & \\ & & & & \gamma_{d+n} & \end{pmatrix} \tag{10}$$

$$a_{n(n+1)} = \begin{pmatrix} (c-d-n)\theta & & & & \\ & \ddots & & & \\ & & & & \\ & & & & (c-d-n)\theta \end{pmatrix} \tag{11}$$

3)  $C_x$  表示系统水平经过一步转移后由  $x$  增加到  $x+1$  的子阵。

当  $x=0$  时, 系统为空,  $C_0$  退化为一个数值。

$$C_0 = \lambda \tag{12}$$

当  $1 \leq x \leq d+1$  时, 全部的匿名用户都在接受虚拟机的服务, 即系统中匿名用户的数量与为匿名用户提供服务 (包括免费服务及注册服务) 的虚拟机数量相等。  $C_x$  为  $(x+1) \times (x+1)$  维方阵, 形式如式(13)所示。

$$C_x = \begin{pmatrix} \lambda & & & \\ & \ddots & & \\ & & & \lambda \end{pmatrix} \tag{13}$$

当  $d+2 \leq x \leq c$  时, 处于活跃状态的虚拟机的数量至少为  $d+1$ , 至多为系统中匿名用户的数量  $x$ 。  $C_x$  为  $\frac{1}{2}(x+d+3)(x-d) \times \frac{1}{2}(x+d+3)(x-d)$  维方阵。  $C_x$  表示成  $\lambda$  倍的单位矩阵。

进一步分析可知, 当  $x > c+1$  时,  $B_x = B_{c+1}$ 。当  $x > c$  时,  $A_x = A_c$ ,  $C_x = C_c$ 。为了表述方便, 令  $B_{c+1} = B$ ,  $A_c = A$ ,  $C_c = C$ , 系统转移率矩阵  $Q$  可表示为分块三对角形式。

$$Q = \begin{pmatrix} A_0 & C_0 & & & & \\ B_1 & A_1 & C_1 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & B_{c-1} & A_{c-1} & C_{c-1} \\ & & & & B_c & A & C \\ & & & & & B & A & C \\ & & & & & & \ddots & \ddots & \ddots \end{pmatrix}$$

转移率矩阵  $Q$  的结构表明, 状态的转移只发生在相邻的系统水平之间。因此, 三维连续时间马尔可夫链  $\{(N(t), Y(t), S(t)), t \geq 0\}$  是一种拟生灭过程, 该过程正常返的充分必要条件是式(14)所示的矩阵二次方程的最小非负解  $R$  的谱半径



表 1 实验参数

参数	设定值
云架构中虚拟机的数量 $c$ /个	10
免费服务率 $\mu_1$ /s <sup>-1</sup>	0.5
注册服务率 $\mu_2$ /s <sup>-1</sup>	0.2
单位时间内虚拟机向匿名用户提供免费服务的能量消耗 $J_{a1}$ /mW	5
单位时间内虚拟机向匿名用户提供注册服务的能量消耗 $J_{a2}$ /mW	4
单位时间内虚拟机处于休眠状态的能量消耗 $J_s$ /mW	0.5
虚拟机由休眠状态转换到活跃状态所产生的能量消耗 $J_r$ /mJ	6

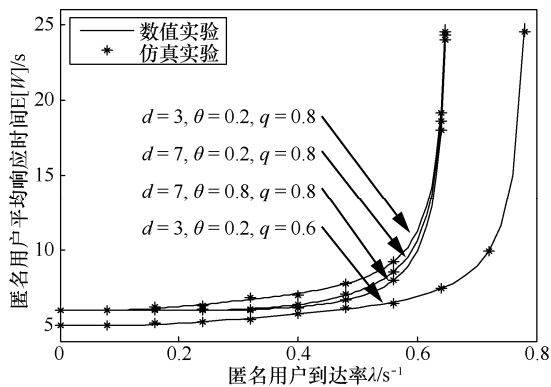


图 2 匿名用户平均响应时间的变化趋势

横向对比分析图 2 可得以下结论。当匿名用户到达率较小(例如  $d=7, \theta=0.8, q=0.8, \lambda < 0.6$ ) 时, 系统中匿名用户的数量一般少于活跃的虚拟机数量, 大部分匿名用户到达系统后可以立即接受免费服务, 其响应时间主要由服务时间决定, 匿名用户平均响应时间几乎不变。当匿名用户到达率较大(例如  $d=7, \theta=0.8, q=0.8, \lambda > 0.6$ ) 时, 系统中匿名用户的数量一般多于活跃的虚拟机数量, 新到达的匿名用户往往先在缓存中排队, 然后才能接受服务。此时, 匿名用户到达率越大, 缓存中排队等待的匿名用户越多, 匿名用户的等待时间越长, 匿名用户平均响应时间呈上升趋势。

纵向对比分析图 2 可得以下结论。固定阈值  $d=3$  及休眠参数  $\theta=0.2$ , 随着注册概率  $q$  的增加, 将有更多的匿名用户试图升级为 VIP 用户, 使注册服务过程占用更长的时间, 造成匿名用户在缓存中的等待时间变长, 匿名用户平均响应时间呈上升趋势。固定阈值  $d=7$  及注册概率  $q=0.8$ , 休眠参数  $\theta$  越大, 一个休眠周期的长度也就越短, 在缓存中有匿名用户等待的条件下, 虚拟机将更早地进入活跃

状态, 匿名用户平均响应时间呈下降趋势。固定休眠参数  $\theta=0.2$  及注册概率  $q=0.8$ , 随着阈值  $d$  的增加, 会有更多的虚拟机处于活跃状态, 云系统的服务能力增强, 匿名用户缓存中的等待时间越短, 匿名用户平均响应时间呈下降趋势。

图 3 表示在不同匿名用户到达率  $\lambda$  下, 阈值  $d$ 、休眠参数  $\theta$  及注册概率  $q$  对系统节能率  $S$  的影响。

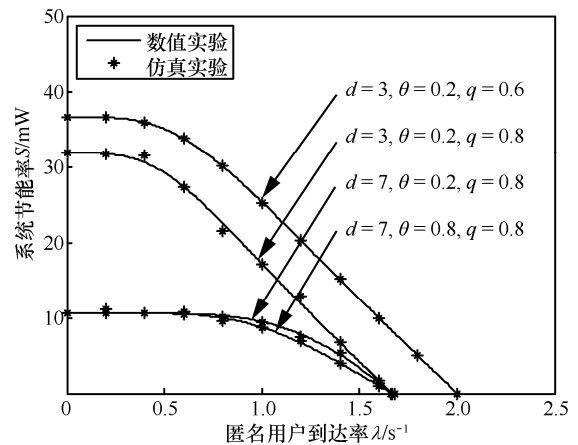


图 3 系统节能率的变化趋势

横向对比分析图 3 可得以下结论。当匿名用户到达率较小(例如  $d=7, \theta=0.8, q=0.8, \lambda < 0.6$ ) 时, 系统中处于活跃状态的虚拟机足以为系统中所有的匿名用户提供服务, 不需要唤醒正在休眠的虚拟机, 系统节能率几乎不变。当匿名用户到达率较大(例如  $d=7, \theta=0.8, q=0.8, \lambda > 0.6$ ) 时, 系统中匿名用户的数量一般要多于处于活跃状态的虚拟机数量, 需要唤醒处于休眠状态的虚拟机以提高云系统的服务能力。匿名用户到达率越大, 需要唤醒的虚拟机数量也就越多, 系统节能率呈下降趋势。

纵向对比分析图 3 可得以下结论。固定阈值  $d=3$  及休眠参数  $\theta=0.2$ , 随着注册概率  $q$  的增加, 将有更多的匿名用户试图升级为 VIP 用户, 为此, 需要更强的云服务能力应付匿名用户的注册请求, 使虚拟机处于休眠状态的可能性减小, 系统节能率呈下降趋势。固定阈值  $d=7$  及注册概率  $q=0.8$ , 休眠参数  $\theta$  越大(休眠周期越短), 在缓存中有匿名用户等待的条件下, 虚拟机将更早地进入活跃状态, 虚拟机处于休眠状态的时间也就越短, 系统节能率呈下降趋势。固定休眠参数  $\theta=0.2$  及注册概率  $q=0.8$ , 随着阈值  $d$  的提高, 会有更多的虚拟机处于活跃状态(更少的虚拟机处于休眠状态), 系统

节能率呈下降趋势。

综上所述，匿名用户到达率越小，匿名用户的平均响应时间越低，系统节能率越高，单个匿名用户对云服务越满意。匿名用户到达率越大，匿名用户的平均响应时间越高，系统节能率越低，即接受云服务的匿名用户也就越多，对云供应商越有利。综合考虑匿名用户访问云服务所获得的收益与等待云服务所消耗的时间成本，研究匿名用户纳什均衡行为与社会最优行为。

### 5 纳什均衡与社会最优

在本文所提云架构中，所有的匿名用户均以最大化个人收益为目的独立地做出是否访问云系统的决定。换句话说，匿名用户的纳什均衡行为是主观的。然而，在设计及实施云系统的过程中，应考虑所有成员以实现社会收益最大化。从匿名用户到达率的角度优化系统性能。遵循 Naor 模型<sup>[20]</sup>的假设条件，针对不可见系统研究匿名用户的纳什均衡与社会最优行为。

1) 令  $X_1$  表示一个匿名用户完成免费服务获得的收益， $X_2$  表示一个匿名用户完成注册服务获得的收益。

2) 每个匿名用户的收益相同，且所有匿名用户的收益可以叠加。

将匿名用户的个人收益定义为一个匿名用户访问云服务所获得的收益与等待云服务所耗费的时间成本之差。匿名用户的个人收益  $G_{ind}(\lambda)$  为

$$G_{ind}(\lambda) = X_1 + qX_2 - g_w E[W] \quad (19)$$

其中， $g_w$  表示匿名用户在系统中逗留单位时间所耗费的成本。

为了确保系统处于稳定状态，设置匿名用户到达率的上界为  $\lambda_{max}$ 。在条件约束范围  $[0, \lambda_{max}]$  内分析匿名用户的纳什均衡到达率。

1) 当  $G_{ind}(0) \leq 0$  时，即使没有新的匿名用户加入缓存，缓存中已有匿名用户的个人收益也均为非正值。匿名用户的纳什均衡到达率为  $\lambda_e = 0$ 。

2) 当  $G_{ind}(\lambda_{max}) \geq 0$  时，即使所有的匿名用户都加入缓存，每个匿名用户的收益也均为非负值。匿名用户的纳什均衡到达率为  $\lambda_e = \lambda_{max}$ 。

3) 当  $G_{ind}(\lambda_{max}) < G_{ind}(\lambda) < G_{ind}(0)$  时，只有部分匿名加入缓存时，才能保证  $G_{ind}(\lambda) \geq 0$ 。当匿名用户个人收益为零时，系统达到纳什均衡状态。此

时，匿名用户的纳什均衡到达率为  $0 < \lambda_e < \lambda_{max}$ 。

将社会收益定义为所有匿名用户的个人收益与系统的节能率之和。社会收益  $G_{soc}(\lambda)$  表示为

$$G_{soc}(\lambda) = \lambda(X_1 + qX_2 - g_w E[W]) + g_s S \quad (20)$$

其中， $g_s$  为系统节能率的影响因子。

通过最大化式(20)给出的社会收益  $G_{soc}(\lambda)$ ，可以得出匿名用户的社会最优到达率  $\lambda^*$ ，如式(21)所示。

$$\lambda^* = \arg \max_{0 \leq \lambda \leq \lambda_{max}} \{G_{soc}(\lambda)\} \quad (21)$$

其中， $\arg \max$  表示使社会收益  $G_{soc}(\lambda)$  取得最大值所对应的匿名用户到达率  $\lambda^*$ 。

为了揭示匿名用户个人收益  $G_{ind}(\lambda)$  与社会收益  $G_{soc}(\lambda)$  的变化规律，下面进行数值实验。采用表 1 给出的实验参数，并设定  $X_1 = 50$ ， $X_2 = 80$ ， $g_w = 10$ ， $g_s = 0.5$ ，匿名用户的个人收益  $G_{ind}(\lambda)$  与社会收益  $G_{soc}(\lambda)$  随匿名用户到达率的变化趋势分别如图 4 和图 5 所示。

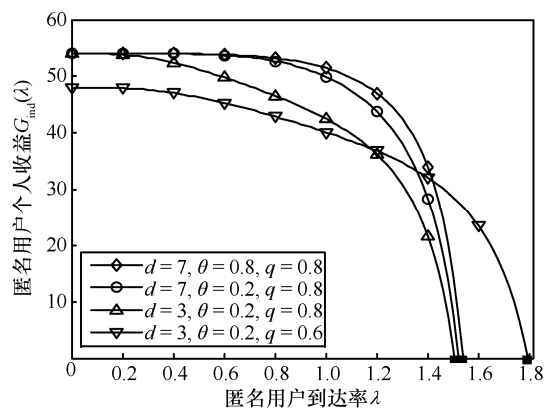


图 4 匿名用户个人收入的变化趋势

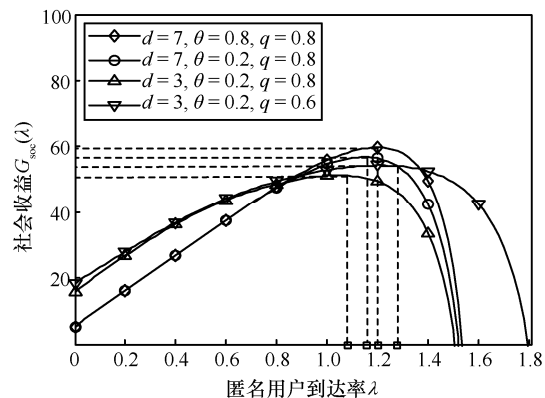


图 5 社会收入的变化趋势

由图 4 可知，对于全部的参数组合，随着匿名

用户到达率的增加，匿名用户个人收益均呈下降趋势。每一个匿名用户都想要进入系统，通过访问云服务获得收益。但是，随着匿名用户到达率的增加，匿名用户平均响应时间增加，导致匿名用户个人收益减小。对于图 4 中的每一条曲线，都存在唯一的匿名用户到达率  $\lambda$  使匿名用户个人收益  $G_{ind}(\lambda) = 0$ ，该到达率即为匿名用户纳什均衡到达率  $\lambda_e$ ，用■进行标注。

由图 5 可知，对于全部的参数组合，随着匿名用户到达率的增加，社会收益均呈现出先上升再下降的趋势。当匿名用户到达率较小时，影响社会收益的主要因素是访问云服务的匿名用户的数量，因此，匿名用户到达率越大，社会收益越大。当匿名用户到达率较大时，匿名用户平均响应时间急剧增加，系统节能率也不断降低，因此，随着匿名用户到达率的增加，社会收益不断降低。对于图 5 中的每一条曲线，都存在唯一的匿名用户到达率  $\lambda$  使社会收益  $G_{soc}(\lambda)$  取得最大值，该到达率即为匿名用户社会最优到达率  $\lambda^*$ ，用□进行标注。

为了比较社会最优到达率  $\lambda^*$  与纳什均衡到达率  $\lambda_e$  的关系，以虚拟机总数  $c = 10$  为例，设置阈值  $d = 1, 3, 5, 7$ ，针对不同的休眠参数  $\theta$  与不同的注册概率  $q$  进行了数值实验。部分实验结果如表 2 所示。

由表 2 可知，首先，对于所有的阈值  $d$ 、注册概率  $q$  及休眠参数  $\theta$ ，匿名用户的社会最优到达率  $\lambda^*$  小于纳什均衡到达率  $\lambda_e$ ，即在只考虑个人收益的情况下，会有更多的匿名用户加入缓存。其次，休眠机制对纳什均衡到达率  $\lambda_e$  及社会最优到达率  $\lambda^*$  的影响如下。固定阈值  $d$  及注册概率  $q$ ，随着休眠参数  $\theta$  的增加，匿名用户的纳什均衡到达率  $\lambda_e$  与社会最优到达率  $\lambda^*$  均呈递增的趋势。针对当前设置的匿名用户完成免费服务获得的收益  $X_1$ 、匿名用户完成注册服务获得的收益及系统节能率的影响因子  $g_w$ ，休眠参数  $\theta$  的取值越高，用户响应性能受节能机制的影响越小，用户访问云系统的积极性越高，匿名用户的纳什均衡到达率  $\lambda_e$  与社会最优到达率  $\lambda^*$  均有所上升。

云供应商可以利用匿名用户的社会最优到达率  $\lambda^*$  与纳什均衡到达率  $\lambda_e$  的差值向匿名用户投放公益广告、进行企业宣传等，在调节匿名用户到达率，实现社会最优的同时，促进企业的良好运行与长久发展。

## 6 结束语

为了应对激烈的市场竞争，云供应商经常要提供免费服务以吸引更多的用户，并努力使匿名用户注册为 VIP 用户。兼顾匿名用户的响应性能与系统节能效果，本文提出了一种带有休眠模式的新型云

表 2 纳什均衡到达率  $\lambda_e$  与社会最优到达率  $\lambda^*$  的数值结果

阈值 $d$	注册概率 $q$	休眠参数 $\theta$	纳什均衡到达率 $\lambda_e$	社会最优到达率 $\lambda^*$	阈值 $d$	注册概率 $q$	休眠参数 $\theta$	纳什均衡到达率 $\lambda_e$	社会最优到达率 $\lambda^*$
1	0.1	0.1	3.300 5	1.536 6	5	0.5	0.1	1.922 6	1.203 1
1	0.1	0.2	3.569 3	2.360 6	5	0.5	0.2	1.989 9	1.357 4
1	0.1	0.3	3.626 1	2.591 4	5	0.5	0.3	2.006 7	1.449 0
1	0.1	0.4	3.650 0	2.696 7	5	0.5	0.4	2.012 4	1.472 2
1	0.1	0.5	3.665 0	2.759 0	5	0.5	0.5	2.016 1	1.501 4
1	0.1	0.6	3.675 4	2.805 6	5	0.5	0.6	2.018 6	1.511 1
1	0.1	0.7	3.683 0	2.850 4	5	0.5	0.7	2.020 5	1.543 6
1	0.1	0.8	3.688 8	2.866 5	5	0.5	0.8	2.022 0	1.548 5
1	0.1	0.9	3.693 5	2.888 1	5	0.5	0.9	2.023 2	1.551 3
3	0.3	0.1	2.402 7	1.267 5	7	0.7	0.1	1.621 7	1.162 6
3	0.3	0.2	2.533 0	1.638 3	7	0.7	0.2	1.640 0	1.240 8
3	0.3	0.3	2.568 1	1.806 2	7	0.7	0.3	1.647 4	1.259 8
3	0.3	0.4	2.587 1	1.870 4	7	0.7	0.4	1.651 6	1.262 0
3	0.3	0.5	2.599 2	1.928 0	7	0.7	0.5	1.654 3	1.289 7
3	0.3	0.6	2.603 3	1.965 4	7	0.7	0.6	1.656 3	1.291 7
3	0.3	0.7	2.606 0	1.951 8	7	0.7	0.7	1.657 7	1.292 4
3	0.3	0.8	2.608 0	1.986 7	7	0.7	0.8	1.658 8	1.304 2
3	0.3	0.9	2.609 6	2.009 7	7	0.7	0.9	1.659 7	1.309 4

架构,通过构建带有二次可选服务且部分虚拟机异步多重休假的排队模型,分析了匿名用户平均响应时间与系统节能率2个性能指标。理论实验及仿真实验显示了阈值、休眠参数及注册概率对系统性能的影响。从经济学的角度出发,研究匿名用户的个人收益与社会收益,揭示出匿名用户纳什均衡到达率要高于匿名用户社会最优到达率。基于此,云供应商可以适度投放广告等,调节匿名用户到达率并实现云系统的社会最优。

### 参考文献:

- [1] KALRA S, SOOD S. ECC-based anti-phishing protocol for cloud computing services[J]. *International Journal of Security and Networks*, 2013, 8(3): 130-138.
- [2] 王斌,李伟民,盛津芳,等. TCCL: 安全高效的拓展云桌面架构[J]. *通信学报*, 2017, 38(Z1): 9-18.  
WANG B, LI W M, SHENG J F, et al. TCCL: secure and efficient development of desktop cloud structure[J]. *Journal on Communications*, 2017, 38(Z1): 9-18.
- [3] 赵辉,丁鸣,程青松,等. SDN与NFV技术在云数据中心的规模应用[J]. *电信科学*, 2016, 32(1): 144-151.  
ZHAO H, DING M, CHENG Q S, et al. Application of SDN and NFV technology in the cloud data center[J]. *Telecommunications Science*, 2016, 32(1): 144-151.
- [4] 张建标,朱元曦,胡俊,等. 面向云环境的虚拟机可信迁移方案[J]. *网络与信息安全学报*, 2018, 4(1): 6-14.  
ZHANG J B, ZHU Y X, HU J, et al. Scheme of virtual machine trusted migration in cloud environment[J]. *Chinese Journal of Network and Information Security*, 2018, 4(1): 6-14.
- [5] RUAN Z, DONG A. Designing and implementation of a RFID-based VIP customer service information management system[C]//The International Conference on Information Engineering and Applications. 2013: 741-748.
- [6] HUH M, KIM D, KIM E, et al. Secure virtual personal cloud service based on CCN/VPC[C]//International Conference on Consumer Electronics. IEEE, 2012: 642-643.
- [7] KLIAZOVICH D, BOUVRY P, GRANELLI F, et al. Energy consumption optimization in cloud data centers[M]. Wiley-IEEE Press. 2015.
- [8] 金顺福,郝闪闪,王宝帅. 融合双速率和工作休眠的虚拟机调度策略及参数优化[J]. *通信学报*, 2017, 38(12): 10-20.  
JIN S F, HAO S S, WANG B S. Virtual machine scheduling strategy based on dual-speed and work vacation mode and its parameter optimization[J]. *Journal on Communications*, 2017, 38(12): 10-20.
- [9] MADAN K. An M/G/1 queue with second optional service[J]. *Queueing Systems*, 2000, 34(1-4): 37-46.
- [10] GHORBANI-MANDOLAKANI M, RAD M. ML and Bayes estimation in a two-phase tandem queue with a second optional service and random feedback[J]. *Communications in Statistics-Theory and Methods*, 2016, 45(9): 2576-2591.
- [11] YANG D Y, CHEN Y H. Computation and optimization of a working breakdown queue with second optional service[J]. *Journal of Industrial and Production Engineering*, 2018, 35(3): 181-188.
- [12] DOSHI B T. Queueing systems with vacations-A survey[J]. *Queueing Systems*, 1986, 1(1): 29-66.
- [13] 马晓彤,金顺福,刘建平,等. 认知无线网络中的基站节能策略及纳什均衡研究[J]. *通信学报*, 2016, 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*, 2016, 37(7): 172-181.
- [14] JAIN A, JAIN M. Multi server machine repair problem with unreliable server and two types of spares under asynchronous vacation policy[J]. *International Journal of Mathematics in Operational Research*, 2017, 10(3): 1-30.
- [15] CHEN Y, CAI L, WEI C M. A discrete-time Geo/G/1 retrial queue with balking customer, second optional service, Bernoulli vacation and general retrial time[C]//The 7th International Conference on Fuzzy Information and Engineering. 2016: 255-266.
- [16] LI X, LI P, HUANG J, et al. Performance analysis of cloud computing centers serving parallelizable rendering jobs using M/M/c/r queueing systems[C]//IEEE International Conference on Distributed Computing Systems. IEEE, 2017.
- [17] RAJENDRAN V V, SWAMYNATHAN S. Queueing model for improving QoS in cloud service discovery[C]//International Conference on Frontiers in Intelligent Computing: Theory & Applications. 2016.
- [18] LATOUCHE G, RAMASWAMI V, SETHURAMAN J, et al. Matrix-analytic methods in stochastic models[M]. Springer, 2013.
- [19] GREENBAUM A. Iterative methods for solving linear systems[M]. Society for Industrial and Applied Mathematics, 2001.
- [20] NAOR P. The regulation of queue size by levying tolls[J]. *Econometrica*, 1969, 37(1): 15-24.

### [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。

武海星(1993-),女,河北邯郸人,燕山大学硕士生,主要研究方向为云资源管理与纳什均衡。

霍甜甜(1994-),女,河北邯郸人,燕山大学硕士生,主要研究方向为排队理论与博弈论的应用研究。

赵文娟(1994-),女,河北衡水人,燕山大学硕士生,主要研究方向为马尔可夫决策过程与智能优化。