基于服务器集群预留机制的高级云体系架构研究*

高宏卿[a]，任淑霞[a]，种大双[b]

（河南师范大学 a. 计算机与信息技术学院；b. 教育技术系，河南 新乡 453007）

摘　要：云计算数据服务中心的巨大能源消耗成为其发展的绿色挑战，提出了基于能源感知的高级云体系架构，并对此体系架构进行深入研究。提出了基于服务器集群预留机制的动态分配策略，并以电力消耗为例进行建模，通过分析电力消耗的成本函数研究该策略的可行性及能源高效性。实验表明，该策略能有效平衡云用户等待忍耐度和电力消耗成本，在最大限度接收云用户的前提下达到电力消耗最低点。

关键词：高级云体系架构；预留机制；任务背叛；电力消耗；最小成本函数；云仿真模拟器

中图分类号：TP393　　**文献标志码**：A　　**文章编号**：1001-3695(2012)07-2593-04

doi:10.3969/j.issn.1001-3695.2012.07.052

Research of high-level cloud architecture based on cloud server cluster reserve mechanism

GAO Hong-qing[a], REN Shu-xia[a], CHONG Da-shuang[b]

(*a. College of Computer & Information Technology, b. College of Educational Technology, Henan Normal University, Xinxiang Henan* 453007, *China*)

Abstract: The vast energy consumption of cloud computing become the green challenge of its development. This paper proposed the high-level cloud architecture based on energy-aware and carried on a thorough research on it. And this paper put forward the dynamic allocation strategy based on server cluster reserve mechanism. Taking power consumption as example for modeling, it studied the feasibility and energy efficiency of this strategy by analyzing the cost function of the power consumption. Experimental result shows that the proposed strategy can effectively balance the waiting degree of cloud users' tolerance and the power consumption cost and minimize the power consumption at the premise of receiving users as more as possible.

Key words: high-level cloud architecture; reserve mechanism; job defection; power consumption; minimize cost function; CloudSim

0　引言

当前信息领域的热点——云计算带动了整个商业模式的转变，延伸出了新的商业体系[1]，在它宣告低成本提供超级计算时代到来的同时也给云计算用户、产业界及各国政府带来深远影响与变革。云计算作为信息技术产业最深刻的变革，真正实现了信息服务的透明化，并以其灵活、高效、低成本、节能的运作方式成为推动 21 世纪产业绿色发展的重要引擎[2]。然而，据 Gartner 调查[2]，全球 IT 业碳排放量已经占到全球总碳排放量的 2%，约为 3 500 万吨，其中数据中心成为碳排放大户，2020 年全球数据服务中心的碳排放量将达到 2 059 亿吨，成为 ICT 碳足迹中增长最快的因素。因此，数据服务中心的巨大能源消耗，特别是电力消耗是一个亟待解决的问题。

一个典型的云计算数据服务中心包括数千台服务器，其耗电量分为两大类[3]：a）有效电力，即负载消耗电力；b）无效电力，即空闲服务器以及电源系统、冷却系统和照明等消耗。目前已经有很多尽可能减少电力消耗的服务器设计方案，但目前最好的服务器设计方案的空闲服务器电力消耗仍占消耗高峰的 65% 以上。本文提出基于能源感知的高级云体系架构，以研究电力消耗问题为例，针对此体系架构提出基于服务器集群预留机制的动态分配策略并建立相关模型。最后，采用 CloudSim 模拟工具进行仿真实验和性能分析。

1　云计算高级体系架构

1.1　云计算简介

云计算（cloud computing）[4] 是一种将计算和存储任务分配给由大量服务集群所构成的云平台的计算模式和服务模式，或者说是网格计算、分布式计算、并行计算、效用计算、自主计算、网络存储、虚拟化、负载均衡等传统计算机和网络技术发展融合的产物。它在吸收了网格计算、效用计算、自主计算等特点的基础上，整合了互联网“云海”中的“云岛”资源，将用户端的计算集中到“云端”[4]，以服务的形式通过互联网提供给用户。随着云计算的快速发展，它所提供的服务就如商品一样流通，人们使用通过互联网传输的云服务就像使用煤气、水电一样，费用低廉、取用方便。总之，云计算将大量用网络连接的计算资源统一管理和调度，构成一个计算资源池向用户提供按需服务。

云计算的技术体系架构并没有统一的标准，但大部分的云基础构架是由通过数据中心传送的可信赖的服务和创建在服

收稿日期：2011-12-20；**修回日期**：2012-10-30　　**基金项目**：河南省教育科学“十一五”规划资助项目（2009-JKGHAG-0321）.

作者简介：高宏卿（1963-），男，河南洛阳人，教授，博士，主要研究方向为计算机网络、云计算（ghq@htu.cn）；任淑霞（1986-），女，河南新乡人，硕士，主要研究方向为计算机网络、云计算；种大双（1987-），男，四川广元人，硕士研究生，主要研究方向为教育技术理论与应用、云计算.

务器上的不同层次的虚拟化技术组成的。通过分析亚马逊、谷歌、微软等供应商的云技术体系架构,云计算的共用技术体系架构[5]包括物理资源层、资源虚拟层、管理中间层、SOA 构建层等四层。本文在此共用技术体系架构的基础上,提出高级云体系架构,为研究基于服务器集群预留机制的服务器动态分配策略奠定了基础。

1.2 高级云体系架构

如图 1 如示,高级云体系架构由以下四部分组成。

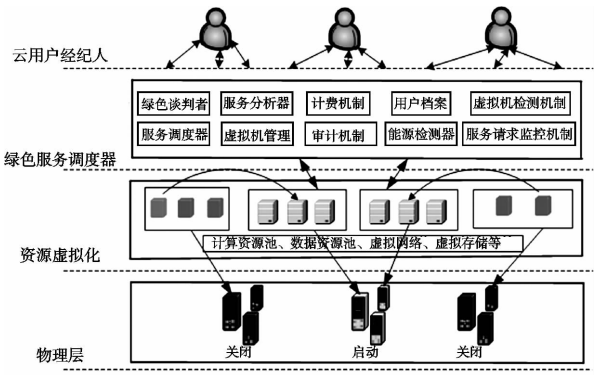

图1 基于能源感知的高级云体系架构

1)云用户/经纪人

云用户即云计算资源的使用者。云用户从终端通过其经纪人向云平台提出自己的要求,包括所需完成任务的 QoS 描述,如 CPU 类型、CPU 数目、内存大小、操作系统及其版本号、成本预算等,而不要预先给出承诺。云用户与部署服务的用户有很大不同,如云用户可以是部署了能够根据用户访问量而实时改变其工作负载的 Web 应用的公司。

经纪人[6,7]代表云用户将世界各地的资源服务请求提交到云中,并快速查找和选择合适其应用需求的资源服务。它支持基于效用函数应用级调度来满足云用户的目标完成服务的分配,仅负责为任务的提交提供调度方案,不负责任务的具体执行。

2)绿色服务调度器

绿色服务调度器是此体系架构的核心模块,是云基础设施与云用户的接口,采用某种谈判机制协调云用户向云平台提交的资源服务请求。绿色服务调度器主要包含以下十种机制:

a)绿色谈判者。从云用户的 QoS 请求与能源节约考虑,它负责与云用户/经纪人进行谈判,最后给出云提供商和云用户之间的具有特定价格和罚款规则的 SLAs 定案。例如,有实验表明,在一个 Web 应用案例中,云供应商在 3 s 内提供的服务可以达到 QoS 标准的 95% 以上。

b)服务分析器。该机制负责翻译并分析首次被提交的服务的需求状况,然后决定是否接受该服务请求,因此该机制必须实时掌握网络、资源等信息。另外,该机制还可以得到资源超载信息,通过“最新状态信息”来关注资源的其他信息,如从虚拟机管理器获得最新的服务器负载信息,从能源监测器获得能源状态信息,从虚拟机监测机制中获得资源可用性信息,从服务请求监测机制中获得工作量处理的信息等,这样有助于更有效地进行资源高效配置。

c)用户档案。收集不同等级的用户特性,赋予重要云用户相应的特权和优先权。

d)计费机制[8]。作为商业交易基础,该机制根据数据管理中心的计算资源的供应和需求,决定服务请求的收费标准及收费模式。例如,可以按服务请求的提交时间(如高峰期/非高峰期)、定价利率(如固定的/变化的)或资源的可用性(如供应/需求)等进行收费。另外,在确定收费标准之前,它需要充分考虑计算资源的供需情况及最大效率的优先服务状况。

e)能源监测器。实时监测虚拟机和物理机能源消耗情况,然后将这些情况提交给虚拟机管理器,供其作出资源高效配置决策。

f)服务调度器。该调度机制为虚拟机分配相应的服务请求,赋予虚拟机资源特权,并命令其执行服务请求。如果云用户提出自动伸缩请求,该机制适时增加或移除虚拟机以满足云用户需求。

g)虚拟机管理器。时刻跟踪虚拟机的可用性以及资源利用情况,负责跨物理机的新虚拟机配置及已有虚拟机的重分配来适应不同虚拟机布局策略。

h)审计机制。监测虚拟机的资源实际使用情况,并计算资源使用成本。另外,对历史记录信息的分析有助于高效分配虚拟机资源。

i)虚拟机监控机制。监视虚拟机的可用性及其资源拥有权。

j)服务请求监控机制。实时监视服务请求的执行进展。

绿色服务调度器通过以上十种机制及其相互作用来扶持高效能的资源管理,并借助机制之间的交互作用来支持面向 SLA 的绿色服务资源管理。

3)资源虚拟化

资源虚拟化包括计算资源池、数据资源池、虚拟网络、虚拟存储等。虚拟化技术[9]是云计算的核心技术之一。根据请求队列,多虚拟机可以动态地在单一物理机上自行启动或停止,将同一物理机上的资源的不同分区配置到服务请求的不同需求上,这为应用程序的运行提供了最大限度的灵活性。此外,多虚拟机还可以在单一物理机上的不同操作系统环境下并发运行多个应用程序。

4)物理层

它是体系架构的最底层,提供了硬件基础设施。通过创建虚拟化资源来满足各种云服务需求,它包含了多个物理机资源,如物理机、服务器、网络设备、存储设备、数据库等。

2 基于服务器集群预留机制的动态分配策略

云计算数据服务中心向云用户提供了各种服务,但背后却需要大量电力来维持其正常运行,换句话说,出乎意料的电力消耗成为阻碍云计算数据服务中心发展的重要因素。有研究提出可以通过关闭空闲服务器达到减少电力消耗的目的,实践证明此方法不可行:由于空闲服务器的不确定性,倘若关闭大量分散服务器的同时突遇任务请求高峰期,便无法及时地、有效地应对。本文在确保服务器提供高性能服务的前提下,提出一种新的基于服务器集群预留机制的服务器动态分配策略,并对此策略进行建模,为研究数据服务中心成本函数打下基础。

2.1 思想概述

本文提出的模型中,数据服务中心[10]两大服务器集群为:a)永久性运行服务器,构成服务主模块(service main module,SMM);b)由任务请求数及其他相关因素决定其是否启动的服务器,构成服务预留模块(service reserved module,SRM)。云用

户任务请求等待服务器响应是具有一定忍耐度的,若等待时间过长,云用户就会离开,此行为称为“任务背叛”,它的发生意味着数据服务中心失去了用户,致使数据中心只产生相应的运行成本却没有收益。任务背叛发生率与处于运行状态服务器数量成反比,而处于运行状态的服务器数量却与电力消耗成正比。所以,任务背叛与电力消耗之间存在一种权衡,其中任务背叛反映了云用户的利益,而电力消耗反映了数据服务中心的利益。故如何设定 SMM 服务器数量成为用户背叛与电力消耗之间最佳平衡状态的关键问题。

图 2 为基于服务器集群预留机制的云数据服务中心示意图。从数据服务中心分离出一定数量的服务器构成 SRM,此模块的状态受上阈值与下阈值两个阈值影响,即当任务请求数超过上阈值时,SRM 开始启动;当任务请求递减到下阈值时,SRM 进入休眠状态。值得注意的是,SRM 的启动过程中消耗电力的同时并没有提供任务服务,因此不能忽略其对数据服务中心整体电力消耗的影响,而其关闭过程可以认为是瞬时完成的。SRM 服务器数量及上、下阈值严格依赖于模拟系统的其他参数以及提出的最小成本函数(minimize cost function, MCF)。其中,最小成本函数由两部分组成,即由于任务背叛产生的成本和服务器的运行成本。

2.2　模型建立

云数据服务中心服务器总数为 N,其中 $n(1\leqslant n\leqslant N)$ 个服务器构成 SMM,其余 $N-n$ 个服务器构成 SRM。每个任务请求自带有独立计时器,它们的到达服从参数 λ 的泊松分布,进入队列后开始计时,计时器的时间呈参数为 $1/\lambda$(即平均计时时间为 λ)的指数分布。当某任务请求在其计时器时间内没有得到服务响应时,发生任务背叛。服务器的服务时间服从参数为 $1/\mu$(即平均服务时间为 μ)的指数分布。

上、下阈值分别用 U、$D(0\leqslant D\leqslant U)$ 表示,共同控制着 SRM 的状态:a)当 SRM 处于关闭状态并且任务请求队列从 U 增到 $U+1$ 时,SRM 开始启动,经过参数为 $1/\nu$ 的指数分布的时间间隔后,SRM 所有服务器均处于运行状态。值得注意的是,此启动过程中 SRM 消耗电力的同时没有提供服务;b)当 SRM 处于运行状态(包括空闲运行与繁忙运行两种),任务请求从 $D+1$ 减到 D 时,SRM 开始关闭,此模型认为关闭过程是瞬时完成的,可以认为几乎没有消耗电力。SRM 的关闭会产生服务中断,中断的任务会立即转向 SMM 的空闲服务器,但若此刻没有空闲服务器,便会记下中断点并排在任务请求队列的队首,并开始时间为 $1/\lambda$(即平均计时时间为 λ)的指数分布的时间计时,等再次得到响应时在中断点处恢复服务。

t 时刻系统状态用二维数组$[I(t),J(t)]$表示。其中,$J(t)$ 表示任务请求队列的当前队长($J(t)=0,1,2,\cdots$)。$I(t)$ 表示 SRM 当前状态,包括三种:a)当 SRM 处于关闭状态时,$I(t)$ 值为 0;b)当 SRM 正在启动或者空闲运行状态时,$I(t)$ 值为 1;c)当 SRM 处于繁忙运行状态时,$I(t)$ 值为 2。根据上面的数组描述,二维数组$[I(t),J(t)]$其实是一个马尔可夫决策过程。马尔可夫决策过程是一类随机过程,此模型中任务请求队列的任务背叛率与任务请求队长有关,根据马尔可夫过程理论,任务背叛具有遍历性。图 3 为决策过程平衡状态的转换图,其概率分布表示为

$$p_{i,j}=\lim_{t\to\infty}P[I(t)=i,J(t)=j]\quad i=0,1,2;j=0,1,\cdots \tag{1}$$

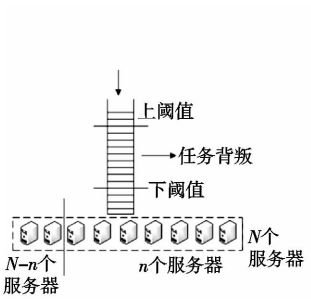

图2　基于服务器集群预留机制的云数据服务中心

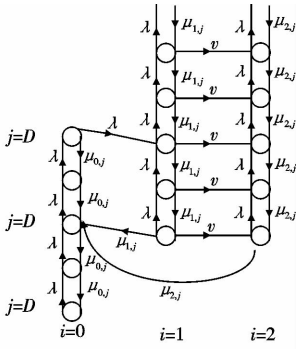

图3　平衡状态转换

图 3 中的 $\mu_{i,j}$ 表示任务请求队列的任务背叛率,随着任务请求队列的队长及繁忙运行服务器数量的变化而变化,表示为

$$\mu_{i,j}=\begin{cases}\min(j,n)\mu+\max(j-n,0)\gamma & i=0,1\\ \min(j,N)\mu+\max(j-N,0)\gamma & i=2\end{cases} \tag{2}$$

图 3 的转换过程叫做拟生灭过程(quasi-birth-and-death process)。拟生灭过程的特点是:一维状态空间除外,其状态空间均是有限的,并且变量跨度为 1(如本模型中任务请求的到达与离开都是独立的)。由于任务背叛率与任务请求队列队长成正比,并且任务请求队列是时刻变化的,所以本模型无法使用矩阵几何和频谱扩展方法。模型假设运行服务器的电力消耗与其是空闲还是繁忙无关,单位时间的平均电力消耗即电力消耗成本函数为

$$C=c_1R+c_2S \tag{3}$$

其中 R、S 是模型中重要的变量,系数及变量说明如下:R 表示由任务背叛引起单位时间内的平均成本;S 表示处于运行状态服务器的运行成本;c_1 表示发生一个任务请求背叛的费用权重;c_2 表示一个服务器处于运行状态的费用权重。

目前还没有相关文献专门研究模型中两个权重系数的关系。若从数据服务中心的实际效益考虑,云提供商希望尽可能多地接收任务请求,进而提高其整体效益,这样会潜意识认为前者比重大于后者比重,即 $c_1>c_2$。事实上,数据中心服务器的成本函数极其复杂,例如考虑到空闲运行服务器的电力消耗是远小于繁忙运行服务器的电力消耗、服务器的关闭过程同样会产生成本。设服务器的启动和关闭的单位成本分别为 d_1、d_2 ($d_1>>d_2$),电力消耗成本函数 C 变为

$$C=c_1R+c_2S_0+c_3S_1+d_1\lambda p_{0,U}+d_2(\mu_{1,D+1}p_{1,D+1}+\mu_{2,D+1}p_{2,D+1}) \tag{4}$$

其中:S_0 表示繁忙运行服务器的平均数量,S_1 表示空闲运行服务器平均数量。为了研究基于服务器预留模块的服务器动态分配策略,本文只对简单的成本函数进行讨论。在讨论成本函数之前需要确定式(1)的稳定状态概率分布,此过程也是解决平衡和规范方程相关设置的必需步骤。

a)当 $i=0$ 且 $j\leqslant D$ 时,即 SRM 处于关闭状态并且任务请求队列当前队长不大于下阈值,计算过程是拟生灭的最简单表现形式,其概率用 $p_{0,D}$ 表示为

$$p_{0,j-1}=\frac{\mu_{0,j}}{\lambda p_{0,j}}\quad j=D,D-1,\cdots,1 \tag{5}$$

b)当 $i=0$ 且 $D<j\leqslant U$ 时,即 SRM 处于关闭状态并且任务请求队列当前队长介于上、下阈值之间,稳定状态的平衡式为

$$(\lambda+\mu_{0,U})p_{0,u}=\lambda p_{0,U-1}$$

$$\lambda p_{0,U}+\mu_{0,U-1}p_{0,U-1}=\lambda p_{0,U-2}$$

$$\cdots$$

$$\lambda p_{0,U}+\mu_{0,D+1}p_{0,D+1}=\lambda p_{0,D} \tag{6}$$

以上平衡式依次产生了期望概率的直接表达式 $p_{0,u}$、$p_{0,D}$。

c)当 $i=1$ 且 $j>U$ 时,即 SRM 处于空闲运行状态并且任务请求队列当前队长大于上阈值,此刻 SRM 处于运行状态并没有提供任何服务。为了简化,假设 $U\geqslant n-1$,这样所有状态的服务率均为 $n\mu$。

以任务请求队列队长为 $U+1$ 为例,任务背叛率表示为 $\eta=\mu_{1,U+1}=n\mu+(U+1-n)\gamma$,相应的平衡式为

$$(\lambda+\nu+\eta)p_{1,U+1}=\lambda(p_{0,U}+p_{1,U})+(\eta+\gamma)p_{1,U+2} \tag{7}$$

$$(\lambda+\nu+\eta+\gamma)p_{1,U+2}=\lambda p_{1,U+1}+(\eta+2\gamma)p_{1,U+3} \tag{8}$$

$$(\lambda+\nu+\eta+2\gamma)p_{1,U+3}=\lambda p_{1,U+2}+(\eta+3\gamma)p_{1,U+4} \tag{9}$$

…

为解决以上无限的平衡表达式组,引进生成函数为

$$g_1(z)=\sum_{j=0}^{\infty}p_{1,U+1+j}z^j \tag{10}$$

式(7)~(9)分别与 1、z、z^2 相乘,然后公式相加并进行等价变换,得到 $g_1(z)$ 的一阶普通线性微分方程为

$$g'_1(z)=[\frac{\lambda}{\gamma}+\frac{\eta}{\gamma z}+\frac{\nu}{\gamma(1-z)}]g_1(z)+\frac{\eta p_{1,U+1}-\lambda z p_U}{\gamma z(1-z)} \tag{11}$$

其中:$p_u=p_{0,u}+p_{1,u}$。

转换为等价封闭表达式为

$$g(z)=\frac{1}{\lambda}e^{\frac{\lambda}{\gamma}z}z^{-\frac{\eta}{\gamma}}(1-z)^{-\frac{\nu}{\gamma}}\int_0^z e^{-\frac{\lambda}{\gamma}z}z^{\frac{\eta}{\gamma}-1}(1-z)^{\frac{\nu}{\gamma}-1}\times(\eta p_{1,U+1}-\lambda z p_U)\,dz \tag{12}$$

3 静态分配策略

云数据服务中心服务器总数为 N,与基于服务器集群的动态分配策略不同的是 n 个服务器处于运行状态,其余 $N-n$ 个服务器处于永久关闭状态。通过动态分配策略模型建立及成本函数分析可以看出静态分配策略其实是本模型的特例,即上阈值为无穷大($U=\infty$)。

设 p_j 为当前任务数为 j 的固定背叛概率,则生成函数为

$$g(z)=\sum_{j=0}^{\infty}p_{n+j}z^j \tag{13}$$

它满足式(11),式(11)转换为

$$g'(z)=[\frac{\lambda}{\gamma}-\frac{n\mu}{\gamma z}]g(z)+\frac{n\mu}{\gamma z}p_n \tag{14}$$

故式(13)可以转换为

$$g(z)=\frac{n\mu}{\gamma}p_n e^{\frac{\lambda}{\gamma}z}z^{-\frac{n\mu}{\gamma}}\int_0^z e^{-\frac{\lambda}{\gamma}z}z^{\frac{n\mu}{\gamma}-1}dz \tag{15}$$

当 $z\to\lambda z/\gamma$ 时,式(15)变为

$$g(z)=\frac{n\mu}{\gamma}p_n e^{\frac{\lambda}{\gamma}z}z^{-\frac{n\mu}{\gamma}}(\frac{\gamma}{\lambda})^{\frac{n\mu}{\gamma}-1}\Gamma(\frac{\lambda}{\gamma}z,\frac{n\mu}{\gamma}) \tag{16}$$

其中:$\Gamma(z,y)$ 是不完全 γ 函数,即

$$\Gamma(z,y)=\int_0^z e^{-x}x^{y-1}dx \tag{17}$$

当 $j<n$ 时的概率 p_j 可以从式(5)中得到。设 $p_n=1$,然后对其进行规格化,则 $g(1)$ 的值为

$$g(1)=\frac{n\mu}{\gamma}p_n e^{\frac{\lambda}{\gamma}}(\frac{\gamma}{\lambda})^{\frac{n\mu}{\gamma}-1}\Gamma(\frac{\lambda}{\gamma},\frac{n\mu}{\gamma}) \tag{18}$$

背叛率 R 即 $\gamma g'(1)$ 为

$$R=\gamma g'(1)=(\lambda-n\mu)g(1)+n\mu p_n \tag{19}$$

4 仿真实验和性能分析

为了验证基于服务集群预留机制的服务器动态分配策略的可行性和能源高效性,本文借助 CloudSim 模拟工具进行相关数值仿真实验。

4.1 CloudSim 简介[11]

CloudSim 是澳大利亚墨尔本大学的网格实验室和 Gridbus 项目宣布推出的继承了 GridSim 的编程模型云计算仿真软件。它体现了云计算最显著的特点,即云计算采用了成熟的虚拟化技术,其扩展部分实现的一系列接口,将数据中心的资源虚拟化为资源池,支持资源监测、资源管理和资源调度模拟。基于数据中心的虚拟化技术、虚拟化云的建模和仿真功能。

4.2 实验与分析

服务器静态分配策略:N 个服务器,其中 n 个服务器组成 SMM,永久处于运行状态,$N-n$ 个服务预留模块处于永久关闭状态。

基于服务预留模块的服务器动态分配策略:N 个服务器,其中 n 个服务器组成 SMM,永久处于运行状态,$N-n$ 个服务器作为 SRM。

实验中的模拟服务器总数为 20 个,即 $N=20$;取平均服务时间作为时间单元,故 $\mu=1$;假设云用户任务请求等待忍耐度为两个时间单位,则 $\gamma=1/2$;任务请求背叛的系数权重为服务器运行的系数权重的 3 倍,故 $c_1=3$,$c_2=1$;λ 取 14(前期实验验证 $\lambda=14$ 表示系统任务负载偏高,为 70%)。

从图 4 中可以看出,RMM 服务器数量为 15,即 $n=15$ 时,两种策略所产生的模拟成本达到最低点。并且当 $n>15$ 时,二者的电力消耗成本几乎一样;但当 $n<15$ 时,即服务器供求处于紧张状态,基于服务预留模块的服务器动态分配策略的电力消耗成本明显低于一般的服务器分配策略,这说明了基于服务预留模块的服务器动态分配策略具有更好的伸缩性与适应性,不但能更好地应对任务请求的突变,更能减少电力消耗,达到减低服务成本的目的。

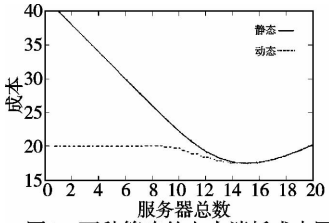

图4 两种策略的电力消耗成本图

5 结束语

云计算时代是一个全新的计算及商业时代。本文从节能的角度提出了基于能源感知的高级云体系架构,并针对此架构提出基于服务预留模块的服务器动态分配策略。仿真实验证明,该策略具有较好的伸缩性与适应性,任务请求较为稳定时该策略与静态分配策略相比各项性能指标相当,但若任务请求突发高峰期时,该策略不但能够及时有效地应对任务请求突变,尽快应对因任务请求突增引起的长时间服务等待,而且其电力消耗成本明显低于现有的静态分配策略,故当任务请求突变前后,该策略的整体性能远远优于静态分配策略。下一步研究重点是服务预留模块的上、下阈值以及其他相关参数对本文提出的基于服务器集群预留机制的动态分配策略的影响。

参考文献:

[1] 孔楠. 基于云计算平台的商业服务模式研究[D]. 上海:上海外国语大学,2011.

(下转第 2623 页)

(上接第 2596 页)

[2] 云计算如何越过低碳关数据中心是关键[EB/OL]. [2010-10-18]. http://server.zol.com.cn/200/2001529_all.html.

[3] JEYARANI R, NAGAVENI N, VASANTH R R. Design and implementation of adaptive power-aware virtual machine provisioner(APA-VMP) using swarm intelligence[J]. Journal of Parallel and Distributed Computing,2012,28(5):811-821.

[4] 刘鹏. 网格计算与云计算[EB/OL]. [2009-03-09]. http://www.chinacloud.cn/download/PPT/GridCloudComputing.ppt.

[5] RODRIGO N C, RAJIV R, ANTON B,*et al*. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms[J]. Software:Practice and Experience,2011,41(1):23-50.

[6] RAJKUMAR B, CHEE S Y, SRIKUMAR V,*et al*. Cloud computing and emerging IT platforms:vision,hype,and reality for delivering computing as the 5th Utility[J]. Future Generation Computer Systems,2009,25(6):599-616.

[7] SAURABH K G, CHEE S Y, ARUM A,*et al*. Environment-conscious scheduling of HPC applications on distributed cloud-oriented data centers[J]. Journal of Parallel and Distributed Computing,2011,71(6): 732-749.

[8] 高宏卿,刑颖. 基于经济学的云资源管理模型研究[J]. 计算机工程与设计,2010,31(19):2-4.

[9] 曾龙海,张博锋,张丽华,等. 基于云计算平台的虚拟集群构建技术研究[J]. 微电子学与计算机, 2010,27(8):1-2.

[10] 杨长兴,吕祯恒. 一种统一的资源预留策略[J]. 计算机工程与应用,2005,41(24):144-146.

[11] CloudSim[EB/OL]. http://baike.baidu.com/view/3652473.htm.