

云计算环境中虚拟资源分配优化策略研究

许力^{1,2}, 曾智斌^{1,2}, 姚川^{1,2}

(1. 福建师范大学 数学与计算机科学学院, 福建 福州 350007;

2. 福建师范大学 网络安全与密码技术福建省高校重点实验室, 福建 福州 350007)

摘要: 针对目前大多数关于数据中心资源分配的研究并没有充分地考虑如何有效地降低能耗, 提出一种高能效的虚拟资源分配策略, 将云计算中的虚拟资源分配问题建模为多目标优化模型, 并使用多目标进化算法求解该模型。仿真结果表明, 针对不同特征虚拟主机和服务器的需求, 策略能够成功地生成调度机制(虚拟资源分配方案), 能有效地降低运行数据中心所需的能耗。

关键词: 云计算; 虚拟资源分配; 绿色计算; 多目标优化; 能量高效

中图分类号: TP 393

文献标识码: A

文章编号: 1000-436X(2012)Z1-0009-08

Study on virtual resource allocation optimization in cloud computing environment

XU Li^{1,2}, ZENG Zhi-bin^{1,2}, YAO Chuan^{1,2}

(1. School of Math and Computer Science, Fujian Normal University, Fuzhou 350007, China;

2. Key Lab of Network Security and Cryptology of Fujian Province University, Fujian Normal University, Fuzhou 350007, China)

Abstract: Most existing researches of resource allocation in data center did not take into full consideration how to decrease energy consumption. The energy efficiency virtual resource allocation for cloud computing as a multi-objective optimization problem was formulated, which was then solved by intelligent optimization algorithm. The simulation results reveal that the strategy can successfully generate schedule scheme of different numbers of servers-VM with diverse characteristics and decrease the total operating energy of data center effectively.

Key words: cloud computing; virtual resource allocation; green computing; multi-objective optimization; energy efficiency

1 引言

云计算被认为是下一代的 IT 服务模式, 受到学术界和工业界的巨大关注。使用云计算, 只需通过计算和存储能力十分有限的终端设备, 就可获得近乎无限的计算能力。在云计算时代, 计算和存储

能力像水电等公用事业一样被组织和管理。一种典型的云计算体系结构^[1] IaaS (infrastructure as a service) 如图 1 所示。社会对云计算需求的不断扩大需要构建规模巨大的数据中心, 而维护其运行需要大量的能量。文献[2]中的研究人员预测到 2020 年, 数据中心能耗将会达到世界总能耗的 2%。在

收稿日期: 2012-08-06

基金项目: 国家自然科学基金资助项目 (61072080); 福建省高校产学研合作科技重大基金资助项目 (2011H6008); 福建省自然科学基金资助项目 (2011J5148); 福建省科技厅 K 类基金资助项目 (JK2011010)

Foundation Items: The National Natural Science Foundation of China (61072080); Industry-University Cooperation Key Project of Science and Technology of Fujian Province (2011H6008); The Natural Science Foundation of Fujian Province (2011J5148); K-class Program of Fujian Provincial Department of Science and Technology (JK2011010)

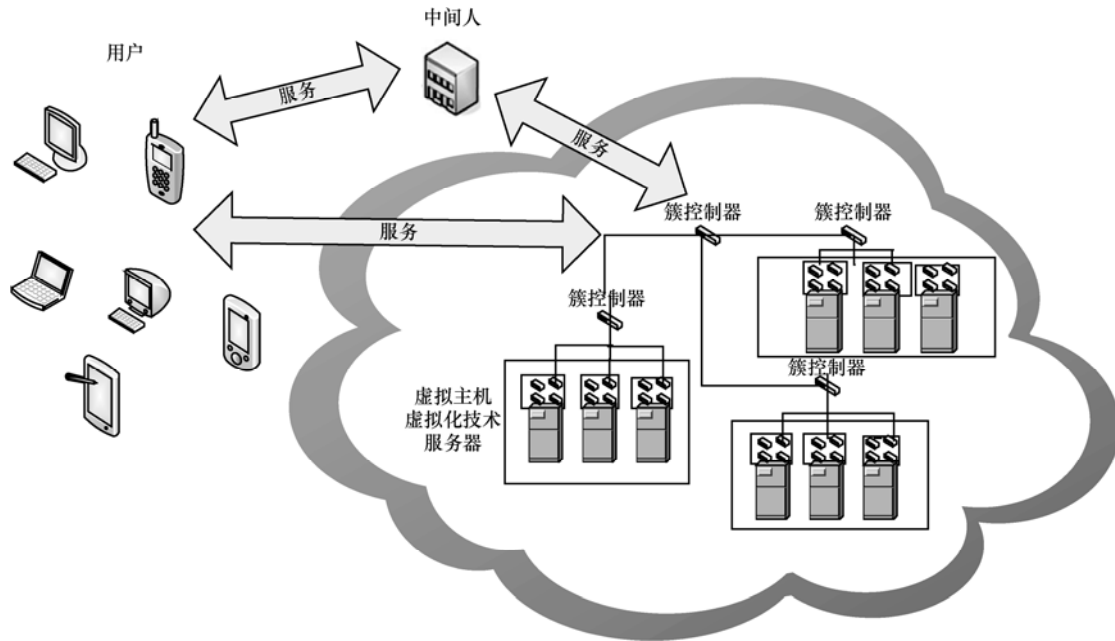


图 1 云计算体系结构

美国，2006 年数据中心消耗了总能耗的 1.5% 并且以每年 18% 的速度增长^[3]。据统计，每使用 1kWh 电将会产生 0.555kg CO₂，这样使能量的供应和环境将会面临前所未有的巨大压力。随着传统能源的逐渐枯竭和价格的飞涨，维护运行数据中心的费用将会超过购买硬件的费用。所以如何高效地运行数据中心是一个亟待解决的问题。

在不考虑散热系统的情况下，数据中心最大的能量消耗是计算过程的能量消耗即动态能量消耗。本文通过虚拟化技术研究如何优化计算过程的能量消耗。虚拟化技术的应用能够之前运行在不同服务器上的虚拟主机聚合在少量的服务器上运行。即使不同的操作系统环境并存于同一物理服务器，也能保持彼此之间以及与底层操作系统之间的隔离。操作系统环境在不同物理服务器之间的迁移可通过动态迁移技术^[4]实现。虚拟化技术使对数据中心的使用更加地灵活、安全、容错、可管理以及经济。然而，目前大部分资源分配的研究主要关注的是通过平衡数据中心的负载而提升整体的性能。这些研究并没有充分地考虑如何降低能耗。而且云计算服务必须满足 SLA (service level agreement)。SLA 是通过用户与供应商之间协商确定的，它规定了严格的性能指标（例如响应时间、吞吐量等）。因此在满足性能指标的前提下如何降低云计算数据中心的能耗带来了巨大的挑战。针对上述问题，本文提出一种高能效的虚拟资源分配策略，将云计算中

的虚拟资源分配问题建模为一个多目标优化模型，并使用多目标进化算法求解该模型。

2 相关工作

高效运行数据中心是资源分配领域的一个新兴课题。针对目前高效运行数据中心所提出的解决方法，根据其作用范围可以分为单服务器级别的管理和服务器集群级别的管理 2 类。前者针对的是单一服务器能耗的优化，而后者针对的是服务器集群的能耗优化。

单服务器级别的管理主要包括研发低能耗的电子元件和动态调压调频技术 DVFS (dynamic voltage and frequency scaling)。其中，DVFS 技术是通过降低 CPU 的频率和供应电压来延长任务的执行周期，从而最大化地消除 CPU 空闲周期来达到节能目标。基于单服务器级别管理的策略^[5,6]是根据负载需求动态改变服务器的 CPU 速率以达到节能。尽管单服务器级别的管理可以有效降低单一服务器的功耗，但无法避免静态能耗即硬件固有能耗的损失，从而无法优化整个集群的能耗。

服务器集群级别的管理主要包括开关服务器和基于虚拟化技术的虚拟资源分配策略。开关服务器是指关闭处于空闲状态的服务器来避免能耗的浪费，当运行的服务器不足以满足当前负载时再开启。其中，文献[7]和文献[8]使用智能模型预测数据中心的负载，从而选择性地关闭冗余服务器以达到

节能。

在如何高效地利用数据中心的研究中, 单服务器级别和服务器集群级别这 2 类管理方法并不是单独使用的, 结合这两者各自优势的研究已经取得了较大的进展。虚拟化是云计算数据中心的一个最基本的特征。虚拟化技术能够在实际的硬件资源上构建虚拟资源, 从而实现硬件资源利用的最大化。如何充分地发挥虚拟化技术的优势, 研究人员提出了不同的节能方案。其中文献[9]和文献[10]将虚拟主机聚合在少数服务器, 但没有充分考虑服务器性能和可承受能力, 因为物理服务器的性能会随着负载的增加而降低。文献[11]中虚拟资源分配问题被建模为多维背包问题, 并找到服务器执行单位事务功耗的最优点, 但并没有提出有效的虚拟资源分配策略。文献[12]是根据对内存的竞争情况调度虚拟主机, 而没有考虑其他资源的竞争。文献[13]根据最大化服务器利用率的思想进行虚拟资源的分配, 但同时导致了过程中大量的迁移操作。文献[10]和文献[14]中尽管实现了高效地调度虚拟主机, 但对 NP 问题的求解和大量的使用排序算法导致复杂度过高。大部分相关研究不能在保证 SLAs 的前提下, 有效地降低数据中心的能耗。此外, 有效地综合单服务器级别和服务器集群级别 2 类节能方法的研究还比较少。

针对上述问题, 本文提出了一种高效的虚拟资源分配策略, 将云计算中的虚拟资源分配问题建模为一个多目标优化模型, 并使用多目标进化算法求解该模型。仿真结果表明, 针对不同特征的虚拟主机和服务器的需求, 策略能够生成调度机制, 从而有效地降低运行数据中心所需能耗。

3 系统模型

3.1 预备知识

本文研究计算过程的能耗优化即 CPU 节能。在不考虑散热系统的情况下, 计算过程的能量消耗是数据中心最大的能量消耗。所以建立 CPU 功耗模型是本文工作的前提。CMOS 电路的功耗由静态功耗和动态功耗 2 部分组成。其中, 静态功耗是 CMOS 电路工作时固有的功率消耗, 所以如何在保证性能的情况下降低动态功耗是最重要的。

动态功耗可由式(1)得到。

$$P = V^2 f C_{EFF} \quad (1)$$

其中, V 、 f 和 C_{EFF} 分别代表供应电压, 频率和交换电容。又因为 f 和 V 成正比关系^[18], 所以 CPU 的动态功耗如式(2)所示。

$$P = \alpha f^3 \quad (2)$$

其中, α 为系数。由于执行完一个特定任务所需要的时间与 CPU 频率成反比, 所以在式(2)的两边同时乘以时间变量 t , 可以得到执行指定任务所消耗的能量 E_{task} 。

$$E_{task} = \alpha \beta f^2 \quad (3)$$

其中, β 等于 f 与 t 的乘积。而且对于一个特定的任务, β 是一个常数。于是

$$E_{task} \propto f^2 \quad (4)$$

式(4)说明降低供应电压可以以二次方的形式降低指定任务的能量消耗。然而, 降低电压会导致任务的执行周期加长。DVFS 技术就是通过降低电压, 延长任务的执行周期来消除 CPU 的空闲周期, 从而有效地降低了能量消耗。这就是 DVFS 技术的基本理论。为了方便起见, 假设数据中心的每台服务器都具有能连续变频的 DVFS 技术。

3.2 相关概念

本文中, $S = \{s_1, s_2, \dots, s_m\}$ 表示云计算数据中心中所有服务器的集合。假设每台服务器 s_j 都具备 DVFS 技术, 最高的运行频率 FS_j^{\max} 和最低的运行频率 FS_j^{\min} 。 FS_j^{\max} 和 FS_j^{\min} 分别是服务器处于最高和最低性能时的运行频率。服务器的运行频率可以在 FS_j^{\max} 和 FS_j^{\min} 之间平稳地改变。 $V = \{v_1, v_2, \dots, v_n\}$ 是被调度到 S 中服务器上运行的虚拟主机的集合。每个 v_i 都有一个为满足最低性能要求的运行频率 FV_i^{\min} , 用于保证 v_i 达到指定的性能。

给定服务器集合 S 和虚拟主机的集合 V , 建立 $n \times m$ 的矩阵 $X_{n \times m}$, 其中, 行和列分别代表虚拟主机和服务器。矩阵 $X_{n \times m}$ 中的元素 x_{ij} 是决策变量, 其定义如下:

$$\begin{cases} x_{ij} = 1, & \text{如果虚拟主机 } v_i \text{ 在服务器 } s_j \text{ 上运行} \\ x_{ij} = 0, & \text{其他} \end{cases} \quad (5)$$

决策变量 f_{ij} 表示 v_i 调度至 s_j 上运行时所占用的频率。本文使用 CPU 频率做为调度因子而不是其他文献(如文献[9]和文献[11])中所使用的内核数量或 CPU 利用率等, 因为在度量不同虚拟主机

对硬件资源的共享情况时, 频率具有更好的可度量性。定义函数 $U_j(X) = U_j(x_{1j}, x_{2j}, \dots, x_{nj})$ 来确定一个服务器是否被使用。

$$\begin{cases} U_j(X) = 0, \text{ 若 } x_{1j}, x_{2j}, \dots, x_{nj} \text{ 都等于 } 0 \\ U_j(X) = 1, \text{ 其他} \end{cases} \quad (6)$$

为了有效地避免在虚拟机的调度过程中, 即虚拟资源分配过程中所产生的服务器开关机代价, 定义 $state_j$ 表示服务器 s_j 的初始状态, 即

$$\begin{cases} state_j = 0, \text{ 初始时服务器 } s_j \text{ 处于关闭状态} \\ state_j = 1, \text{ 初始时服务器 } s_j \text{ 处于运行状态} \end{cases} \quad (7)$$

3.3 建立问题模型

将所要研究的问题建立为一个多限制多目标优化模型, 其数学形式如式(8)所示。

$$\begin{aligned} & \text{Minimize } \sum_{j=1}^m U_j(X), \sum_{j=1}^m \left(\sum_{i=1}^n f_{ij} x_{ij} \right)^3, \\ & \sum_{j=1}^m \left(\sum_{i=1}^n f_{ij} x_{ij} - FS_j^{\max} \right)^2 U_j(X) \text{ 且} \\ & \text{minimize } \sum_{j=1}^m \{ (1 - state_j) U_j(X) COST_j^{\text{on}} + \\ & \quad state_j (1 - U_j(X)) COST_j^{\text{off}} \} \end{aligned} \quad (8)$$

$$\text{其中, } x_{ij} \in \{0, 1\}, \quad i = 1, 2, \dots, n; j = 1, 2, \dots, m \quad (9)$$

$$\sum_{j=1}^m x_{ij} = 1, \quad \forall i \quad (10)$$

$$FV_i^{\min} \leq f_{ij} \leq FS_j^{\max}, \quad \forall i, \forall j \quad (11)$$

$$\sum_{i=1}^n f_{ij} x_{ij} \leq FS_j^{\max}, \quad \forall j \quad (12)$$

式(8)中的目标函数包括 4 个: 1) $\sum_{j=1}^m U_j(X)$ 表示使用尽可能少量的服务器满足运行虚拟机要求的目标。这是最重要的目标, 因为在单服务器级别上将服务器关闭比其他任何的节能技术都有效; 2) $\sum_{j=1}^m \{ (1 - state_j) U_j(X) COST_j^{\text{on}} + state_j (1 - U_j(X)) COST_j^{\text{off}} \}$ 的目标函数代表服务器开关机的代价。 $COST_j^{\text{on}}$ 表示开启服务器 s_j 的代价, $COST_j^{\text{off}}$ 表示关

闭服务器 s_j 的代价。当 $state_j = 0, U_j(X) = 1$ 时, 即服务器需要被开启时, 式(8)的值增加开启代价 $COST_j^{\text{on}}$; 当 $state_j = 0, U_j(X) = 0$ 时, 即服务器不需要被开启时, 式(8)的值增加 0; 当 $state_j = 1, U_j(X) = 0$ 时, 即服务器需要被关闭时, 式(8)的值为 $COST_j^{\text{off}}$; 当 $state_j = 1, U_j(X) = 1$ 时, 即服务器不需要被关闭时, 式(8)的值为 0; 3) $\sum_{j=1}^m \left(\sum_{i=1}^n f_{ij} x_{ij} \right)^3$ 代表降低所有虚拟主机在服务器上消耗的总功耗之和的目标; 4) $\sum_{j=1}^m \left(\sum_{i=1}^n f_{ij} x_{ij} - FS_j^{\max} \right)^2 U_j(X)$ 是基于充分利用已用服务器的原则而设置的目标。

限制函数式(9)作为调度虚拟主机的限制, 当 $x_{ij} = 1$ 时虚拟主机 v_i 被调度至服务器 s_j 上运行。限制函数式(10)限制一个虚拟主机只能被调度到一台服务器上运行。限制函数式(11)是性能限制函数。用于保证虚拟主机的运行频率在能够保证 QOS 的性能之上, 即满足 SLAs 中协商的协议 (比如响应时间等)。同时, 限制函数式(11)也考虑了服务器的负载能力。限制函数式(12)是用于对服务器负载能力的考量。即任意一台服务器上所运行的所有虚拟主机的频率之和不能超出这台服务器的运行频率最大值。

决策变量 x_{ij} 和 f_{ij} 构成了最终的解。 x_{ij} 决定了服务器与虚拟主机之间的映射关系。 f_{ij} 决定虚拟主机所占用服务器频率的值。式(8)中的 4 个目标函数间存在一定的冲突关系, 这个数学形式是一个典型的多目标优化问题。

4 解决方案

一般情况下, 多目标优化模型的目标函数之间存在冲突, 对一个目标函数的优化可能会导致其他目标函数的劣化, 所以同时优化所有的目标函数是不可行的。取而代之的是帕累托最优解集。然而, 帕累托最优解集空间往往是巨大甚至无穷的, 在实际应用中, 是根据实际的需要选取所需要的解。传统解决最优化问题的方法不适用于解决多目标最优化问题。故本文引进 NSGA-II 求解式(8)中建立的多目标优化问题, 它是一种性能优秀的多目标进化算法。本文所提出的虚拟资源分配策略的完整流程如图 2 中伪代码所示。

```

Input: 服务器和虚拟主机数量
Output: 调度机制
Begin
随机产生原始种群 (OP);
在 OP 中不执行非劣效性排列;
Then 在原始种群的基础上, 通过联赛选择方法, 交叉和变异产生新的种群
while 未达到终止条件 do
联合新旧种群产生交配种群
Then 执行非劣效性排序;
    对上一步得到种群按拥挤距离在帕累托前沿面上的进行排序, 选择若干个个体构造新种群;
在上一步的新种群基础上通过复制, 交叉和变异产生新种群;
end
end
在可行解中选择一个最优调度方案作为最终的解决方案;
根据调度方案将虚拟主机调入服务器;
编辑和记录数据中心相关数据。

```

图 2 虚拟资源分配策略的流程

5 多目标优化模型存在的问题

5.1 存在问题描述

3.3 节中所构建的多目标优化模型与传统的多目标优化模型有所区别。在传统的多目标优化模型中目标函数拥有相等的权重。而本文研究的是如何降低数据中心的能耗, 所以第一个目标函数是最重要的且具有最大的权重。基于这个原因, 在进化的过程中人工地加强第一个目标函数的重要性。具体的做法是在种群初始化阶段, 构造大量能使第一个目标函数值较小的可行解, 逻辑意义是将虚拟主机集中在尽可能少的物理服务器上运行。这样的可行解在进化的过程中会被大量地选择和保留, 从而达到人工干预进化的目的。

此外, 式(10)的限制函数作用是使一个虚拟主机只能被调度到一台物理服务器上运行。但由于该限制条件过于苛刻, 导致 NSGA-II 算法在进化过程中难以通过变异来提高遗传算法对于局部解空间的搜索能力。原始的 NSGA-II 算法对二进制编码的染色体段的变异方法为 bitwise mutation, 即对每一位以一定的概率进行从 1→0 或从 0→1 的操作, 通过这样的变异方式得到的新染色体存在极大的可能性不满足限制条件, 即非本模型的可行解。导致在解这个多目标优化问题时, NSGA-II 算法出现早熟的现象。针对上述问题, 本文使用 bitcoordinate mutation 改进 NSGA-II 算法。

5.2 Bitcoordinate mutation 的描述

1) 当需要对染色体中的某一位进行从 0 置为 1 的操作时, 将其余对应于该虚拟主机的染色体位置为 0, 逻辑含义是当某虚拟主机在一台服务器上调度执行时, 不允许其同时运行于其他服务器上, 即满足限制函数式(10), 如图 3 所示。

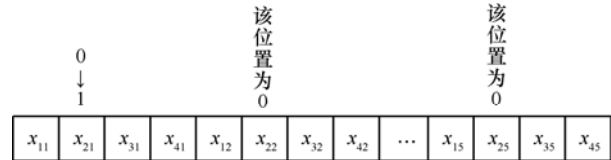


图 3 位协作变异(0→1)

2) 当需要对染色体中的某一位进行从 1 置为 0 的操作时, 从其余对应于该虚拟主机的染色体位上任选一位置为 1, 剩下的置为 0。逻辑含义是当某虚拟主机在一台服务器上关闭时, 从其他的服务器中任选一台调度执行该虚拟主机, 即满足限制函数式(10), 如图 4 所示。

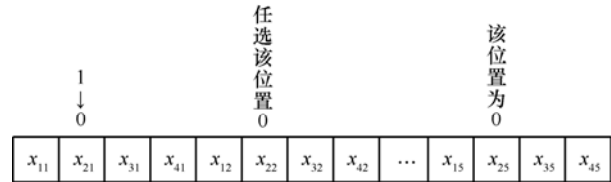


图 4 位协作变异(1→0)

改进后的 bitcoordinate mutation 变异算法, 在种群多样性上表现出优异的性能, 不仅种群的分布比较广, 而且还得到更优的值, 如图 5 所示。

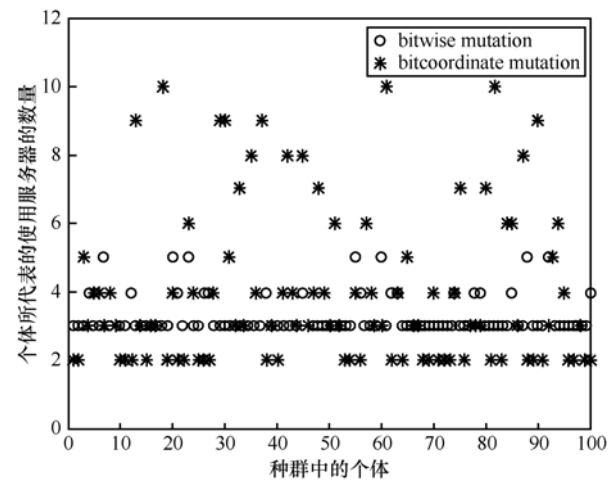


图 5 变异过程得到解的分布情况

6 仿真性能与分析

6.1 仿真环境

仿真平台使用 C++ 开发。平台模拟出不同数量服务器的数据中心，并随机模拟出不同特征的虚拟主机用于调度至服务器上运行。服务器的数量在 20（代表小规模的数据中心）~100（代表中等规模的数据中心）之间变化。值得注意的是管理成千上万台服务器的数据中心是困难的，所以一种可行的解决方案是将其分簇为小规模的集群进行管理从而获得更好的可扩展性。随机赋予每台服务器不同的 CPU 频率：4GHZ、8GHZ、16GHZ。每个虚拟主机的频率可在其所需频率资源的下限和上限之间变化（如：1GHZ~4GHZ）。仿真平台中实现了使用 NSGA-II 的改进，并对 3.3 节所构建的多目标优化模型构造了 4 个目标函数。决策变量 f_{ij} 和 x_{ij} 分别使用实数编码和二进制编码。为了分析仿真实验的结果，做以下重要的假设：1）每台服务器都具备能平稳连续地变化其运行频率的 DVFS 技术；2）CMOS 电路中一个单位的势差可以产生一个单位的 CPU 频率，改变这样的假设数据不会对仿真实验结果产生显著的影响；3）在服务器全负载状态时，静态固有的能量消耗占总能耗的 60%（即静态能耗）。设置 NSGA-II 的最大进化代数为 250 代。对于每一组参数，仿真过程运行数轮后对仿真结果取平均值。表 1 中所示的是仿真过程中 NSGA-II 的一些重要参数及其设置。运行仿真平台的物理服务器硬件配置为 Pentium(R) Dual-Core CPU E5 200, 2.50GHZ, 2G RAM，操作系统：Microsoft Windows XP。

表 1 NSGA-II 的参数及其设置

参数	设置
种群规模	100
交叉概率	0.6
实数编码向量的变异概率	1/RealLength (RealLength 是实数编码向量的长度)
实数编码向量的交叉分布指数	20
实数编码向量的变异分布指数	20
二进制编码向量的变异概率	1/BinaryLength (BinaryLength 是二进制编码向量的长度)

6.2 仿真性能与分析

图 6 所示是在不同规模的数据中心中，用于满足运行虚拟主机的需要而使用的服务器数量的变化情况。从图 6 中可以看出，随着虚拟主机数量的增加，曲面呈平缓上升的趋势。

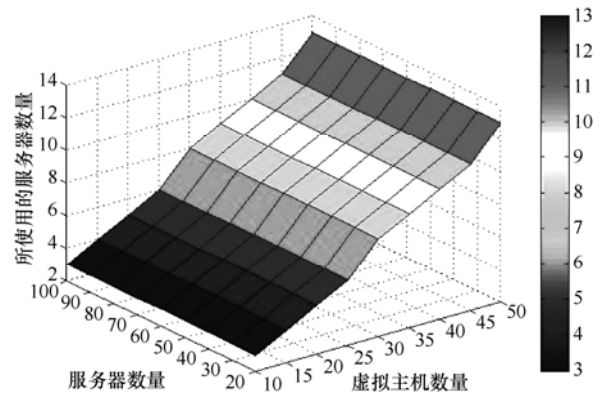


图 6 使用服务器的数量

图 7 所示是虚拟主机在所有服务器上所产生的动态能量消耗之和。从图 7 中可以看出随着虚拟主机数量的增加，代表动态能耗之和的曲面平稳地上升，说明随着虚拟主机和服务器数量变多，即所要解决的问题变复杂时，MOBVRAS 策略仍然可以有效地控制动态能量的总和。图 8 所示是被使用服务器的总体利用率的情况（曲面越低，总体利用率情况越好）。如图 8 所示，曲面呈现上下交替的状态。图 9 所示是数据中心的总能量消耗，由图 9 可知，MOBVRAS 策略能够有效地降低运行数据中心所需能耗。

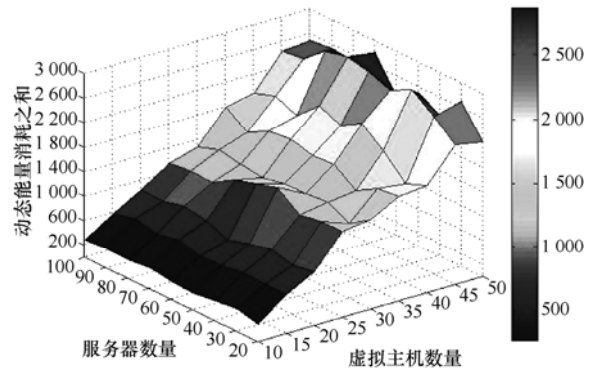


图 7 动态能量消耗之和

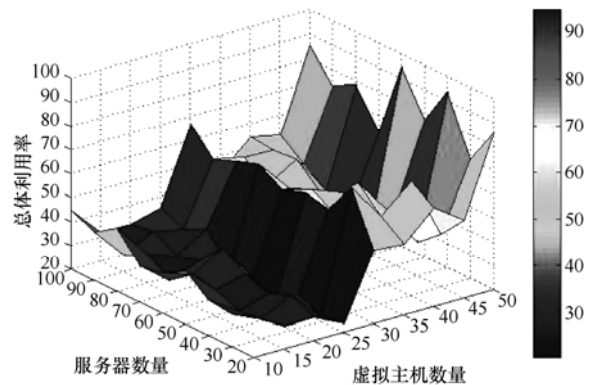


图 8 总体利用率

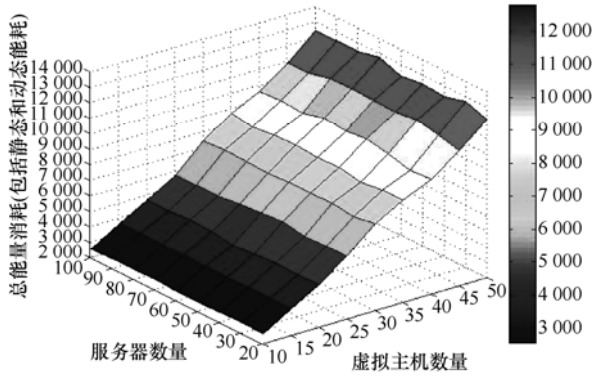


图 9 总能量消耗 (包括静态和动态的能量消耗)

图 10 和图 11 所示分别为将关机代价作为目标函数和未将关机代价作为目标函数时,生成虚拟资源分配方案时所产生的的开关服务器代价,如图 10 和图 11 所示,随着所需要调度执行的虚拟主机的数量增加和数据中心规模的增大,即问题复杂度增高时,所产生的的关机代价也越来越大。但在将关机代价作为目标函数时,MOBVRAS 策略所产生的的服务器关机代价相对较小。保证了在调度执行虚拟主机的过程中,产生较少量的开启和关闭服务器的过程。因为 MOBVRAS 策略的研究重点不包括整个数据中心负载的动态变化情况,所以并没

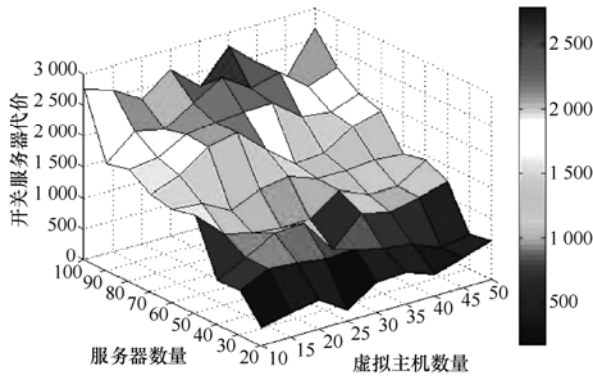


图 10 开关服务器代价 (将开关服务器代价作为目标函数)

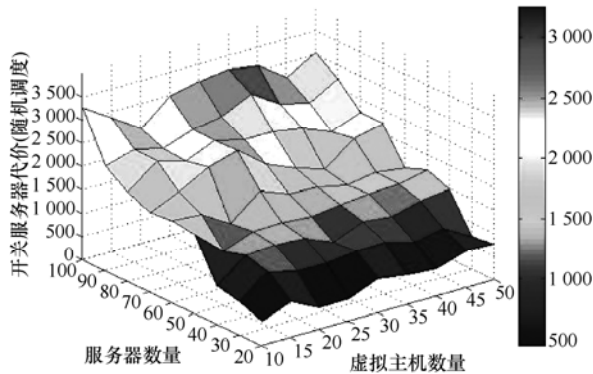


图 11 开关服务器代价 (未将开关服务器代价作为目标函数)

有将减少所产生的的服务器关机代价作为最重要的优化目标,故在减少服务器代价方面的性能并未达到最优。这部分的内容将作为今后工作的一个研究重点。

7 结束语

本文研究如何在保证性能的前提下降低运行数据中心的能耗。提出了一种高能效的虚拟资源分配策略,将云计算中的虚拟资源分配问题建模为一个多目标优化模型,并使用多目标进化算法 NSGA-II 求解该模型。为了表明策略的优越性,开发了仿真平台,并在平台中模拟了不同的条件来测试策略的性能。仿真结果表明,针对不同特征虚拟主机和服务器的需求,策略能够成功地生成虚拟资源分配方案,从而有效地降低运行数据中心的能耗。

在未来的工作中,将引入更优秀的优化算法来求解所构建的多目标优化模型,并考虑服务器的关机代价等更为复杂的因素。所提策略将与该领域的相关研究进行比较分析。为了进一步测试策略的实际可用性,策略将会被使用在诸如 XEN 等虚拟化技术的真实云计算环境中。

参考文献:

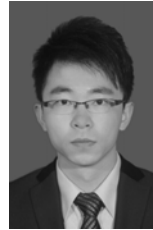
- [1] NURMI D, WOLSKI R, GRZEGORCZYK C, *et al.* The eucalyptus open-source cloud-computing system[A]. 9th IEEE/ACM International Symposium on Cluster Computing and the Grid (CCGRID)[C]. Shanghai, China, 2009. 124-131.
- [2] FORREST W. How to cut data centre carbon emissions?[EB/OL]. <http://www.computerweekly.com/Articles/2008/12/05/233748/how-to-cut-data-centre-carbon-emissions.htm>,2012.
- [3] Report to congress on server and data center energy efficiency [EB/OL].http://www.energystar.gov/ia/partners/Prod_development/downloads,2012.
- [4] CLARK C, FRASER K, HAND S, *et al.* Live migration of virtual machines[A]. Proceedings of the 2nd Conference on Symposium on Networked Systems Design & Implementation[C]. Berkeley, CA, USA, 2005. 273-286.
- [5] TAN Y S, WU Q B, TANG H M. Multi-tier energy management strategy for HPC clusters[A]. Green Computing and Communications (GreenCom)[C]. HangZhou, China, 2010. 112-116.
- [6] KIM K H, BELOGLAZOV A, BUYYA R. Power-aware provisioning of cloud resources for real-time services[A]. Proceedings of the 7th International Workshop on Middleware for Grids, Clouds and e-Science (MGC2009)[C]. Urbana Champaign, USA, 2009.1-6.

- [7] DUY T V T, SATO Y, INOBUCHI Y. Performance evaluation of a green scheduling algorithm for energy savings in cloud computing[A]. Proc 24th IEEE International Parallel and Distributed Processing Symposium (the 6th Workshop on High-Performance, Power-Aware Computing)[C]. Atlanta, GA, USA, 2010.1-8.
- [8] LIU B, YANG J, DIAO Y. Dynamic cluster configuration strategy for energy conservation based on online load prediction[J]. Computer Engineering, 2010,36(24): 96-98.
- [9] YOUNGE A J, LASZEWSKI G V, WANG L, *et al.* Efficient resource management for cloud computing environments[A]. Proceedings of the IEEE International Green Computing Conference (IGCC)[C]. Chicago, USA, 2010.357-364.
- [10] WANG Y, WANG X. Power optimization with performance assurance for multi-tier applications in virtualized data centers[A]. 39th International Conference on Parallel Processing Workshops[C]. San Diego, USA, 2010.1-8.
- [11] SRIKANTIAH S, KANSAL A, ZHAO F. Energy aware consolidation for cloud computing[A]. Proc HotPower'08 USENIX Association, Berkeley, USENIX Association Berkeley[C]. CA, USA, 2008. 10.
- [12] DHIMAN G, MARCHETTI G, ROSING T. VGreen: a system for energy efficient computing in virtualized environments[A]. ISLPED '09 Proceedings of the 14th ACM/IEEE International Symposium on Low Power Electronics and Design[C]. New York, USA, 2009. 243-248.
- [13] LI B, LI J X, HUAI J P, *et al.* Enacloud: an energy-saving application live placement approach for cloud computing environments[A]. CLOUD '09 IEEE International Conference[C]. Bangalore, 2009. 17-24.
- [14] CARDOSA M, KORUPOLU M, SINGH A. Shares and utilities based power consolidation in virtualized server environments[A]. Proc of IFIP/IEEE Integrated Network Management[C]. Long Island, NY, 2009. 327-334.
- [15] AYDIN H, MELHEM R, MOSSE D, *et al.* Dynamic and aggressive scheduling techniques for power-aware real-time systems[EB/OL]. http://www.cs.pitt.edu/PARTS/papers/RTSS01_aydin.pdf, 2012.
- [16] DEB K, PRATAP A, AGARWAL S, *et al.* A fast and elitist multi-objective genetic algorithm: NSGA-II[J]. IEEE Trans on Evolutionary Computation, 2002,6(2):182-197.

作者简介:



许力 (1970-), 男, 福建福州人, 福建师范大学教授、博士生导师, 主要研究方向为物联网与云计算、网络与信息安全、智能信息处理、复杂系统和网络的建模与仿真。



曾智斌 (1985-), 男, 福建莆田人, 福建师范大学硕士生, 主要研究方向为云计算和性能分析、网络与信息安全。



姚川 (1988-), 男, 福建福州人, 福建师范大学硕士生, 主要研究方向为云计算、网络与信息安全等。