

云计算数据中心服务器数量动态配置策略

卫 星^{①②} 张建军^{*①②} 石 雷^① 翟 琰^①

^①(合肥工业大学计算机与信息学院 合肥 230009)

^②(安全关键工业测控技术教育部工程研究中心 合肥 230009)

摘 要: 云计算数据中心由通过高速网络连接的大量服务器构成,一种有效的节能措施是维持与系统负载成比例的活跃服务器数量同时切换剩余服务器到空闲模式,由此分别产生操作能耗和切换能耗。该文研究如何动态配置活跃服务器数量以最小化数据中心能耗(操作与切换能耗之和)的问题。首先,建立了问题的NP数学模型,并分析了无切换能耗情况下最优解的特性;其次,通过消除整数动态规划的递推过程,推导具有多项式复杂度的最优静态算法;最后,采用对未来负载的最坏预测结果作为约束制定了优化在线策略。仿真结果表明,所提出的静态最优和动态优化策略能够适应外界负载的剧烈变化趋势始终谨慎调整活跃服务器和休眠服务器的比例,以接近最优的能耗代价维持数据中心的平稳运行。

关键词: 云计算; 数据中心; 活跃服务器; 离线最优算法; 动态规划; 在线算法

中图分类号: TP393

文献标识码: A

文章编号: 1009-5896(2015)08-2007-07

DOI: 10.11999/JEIT141286

Dynamic Active Servers Allocating Policy for Cloud Computing Data Centers

Wei Xing^{①②} Zhang Jian-jun^{①②} Shi Lei^① Zhai Yan^①

^①(School of Computer and Information, Hefei University of Technology, Hefei 230009, China)

^②(Engineering Research Center of Safety-critical Industry Measure and Control Technology of Ministry of Education, Hefei 230009, China)

Abstract: Cloud computing data centers generally consist of a large number of servers connected via high speed network. One promising approach to saving energy is to maintain enough active servers in proportion to system load, while switch left servers to idle mode whenever possible. Then operating cost and switching cost is brought about respectively. The problem of right-sizing active servers to minimize energy consumption (total cost of operating and switching) in data centers is discussed. Firstly, the NP-hard model is established, and the characteristics of the optimal solution when omitting the switching cost are analyzed. Then by revising the solution procedure carefully, the recursive procedure is successfully eliminated. The optimal static algorithm with polynomial complexity is achieved. Finally, the online strategy is developed using the worst predicting load as the constraints. Simulation results show that the proposed offline and online algorithm can adapt the dramatic trend of external load and always carefully adjust the proportion of active servers, to guarantee minimum power consumption with a smooth computing process.

Key words: Cloud computing; Data center; Active servers; Offline optimal algorithm; Dynamic programming; Online algorithm

1 引言

云计算通过整合存储和计算能力有限的大量终端服务器,使得系统用户只需通过网络“透明”的

访问其中一台服务器就可获得近乎无限的计算能力以及语音、视频、信息搜索等服务,而资源由云计算数据中心统一调度、组织和管理^[1-3]。Amazon, Google, IBM, Microsoft等相继推出以集群计算为模型的云计算数据中心,采用层次结构实现且承载的主要是客户机/服务器模式应用,具有如下典型特征:(1)数据中心内部各服务器间具有高传输带宽。(2)数据中心能够实现服务器和虚拟机的便捷配置和迁移。(3)数据中心支持数十万甚至上百万台的服

2014-10-09收到, 2015-04-16改回, 2015-06-09网络优先出版
国家自然科学基金(61370088), 国家国际科技合作专项项目
(2014DFB10060)和中央高校基本科研业务费专项资金
(2011HGBZ1321, 2012HGQC0012)资助课题

*通信作者: 张建军 jianjun@hfut.edu.cn

务器，并允许增量的部署和扩展，其服务能力远大于外部应用需求。

满足服务需求的同时降低数据中心能耗具有重要的研究意义^[4-6]，即使是以高效能著称的数据中心其资源利用率大部分情况下也仅仅保持在5%~20%，说明绝大部分服务器常处于空闲状态。一种有效的节能措施是动态调整服务器数量以适应于不同时期的负载，与此同时尽量将未被分配到工作进程的服务器切换到休眠模式^[5]。文献[7~13]研究如何通过虚拟技术、负载均衡和负载迁移技术实现大规模服务器集群以及数据中心的绿色节能运营，涉及到分发错误率、用户QoS和SLA保证等，为相似的后续策略研究提供了较好的研究基础。文献[14]研究了相似的最小化内容分发网络能耗问题并给出了优化算法，然而其假设操作能耗函数是一种线性函数，只适合于某些严格实时的特殊网络服务系统，且任意周期的操作能耗与活跃服务器数量无关，简化了问题。文献[15]提出了在服务器集群中采用切换服务器模式，优化能源反应时间(energy-response time product)的指标，并提出了近似最优的贪婪算法。文献[16]给出了负载均衡数据中心通过切换空闲服务器到休眠模式以节约能源的最优算法，然而其认为数据中心的服务器数量巨大而放宽了其必须整数的约束，不符合通常的习惯认知。

本文研究如何动态配置各时隙的活跃服务器数量从而最小化数据中心能耗的问题。首先，从数据中心工作模式出发，将任务分发策略简化为负载均衡方式并建立了问题数学模型；其次，分析了无切换能耗情况下最优解的特性，并给出了平周期与跟随周期递推法则；接下来通过消除整数动态规划的递推过程，给出了具有多项式复杂度的静态最优算法；最后以未来负载的最坏预测结果为约束制定了在线算法。

2 系统模型及问题

设数据中心由 N 台同构的服务器组成，每台服务器可同时支持的服务进程数量为 M 。将数据中心的服务器周期人为划分成相同的离散时隙(例如10 min为单位)，并且考虑一个足够长的时间跨度 $T < \infty$ (例如以月、年为单位)。每一个时隙 $t \in [0, 1, \dots, T]$ 内到达数据中心的负载为 λ_t ，显然 $\lambda_t \leq NM$ ，符合数据中心的运行维护机制则有 $\lambda_0 = \lambda_T = 0$ 。数据中心节能问题是合理配置每一时隙内活跃(开启并提供服务进程)服务器数 $x_t, t \in [0, 1, \dots, T]$ ，以及该时隙内每一活跃服务器承担的负载 $\lambda_{i,t}, i \in [1, x_t]$ ，使得数据中心从时隙0运行到时隙 T 的总能量代价最小。数据

中心的运营代价包括两部分，一是维持各时隙内服务进程所消耗的能量即操作能耗，二是时隙之间活跃服务器数量切换所消耗的能量即切换能耗。

首先推导操作能耗函数 $f(\cdot)$ 。操作能耗一般由延迟成本和能源成本两部分组成。服务器 i 可以视作平均到达率为 $\lambda_{i,t}$ ，平均服务率为 μ 的服务台，由M/GI/1排队模型可知，该服务台平均响应时间是 $1/(\mu - \lambda_{i,t})$ 。令用户能够容忍的响应时长 d_0 、延迟系数 d_1 均为正常数，则服务台 i 单位负载的延迟成本为 $d_1(1/(\mu - \lambda_{i,t}) - d_0)^+$ ，其中符号 $(y)^+$ 表示 y 与0之间的较大值。从而服务器 i 的延迟成本为 $D(\lambda_{i,t}) = \lambda_{i,t}d_1(1/(\mu - \lambda_{i,t}) - d_0)^+$ 。能源成本包括电能分配与消耗。令基本电能消耗 e_0 、消耗系数 e_1 均为正常数，则服务器 i 的能源成本为 $E(\lambda_{i,t}) = e_1\lambda_{i,t} + e_0$ 。

从而操作能耗函数为

$$f(\lambda_{i,t}) = D(\lambda_{i,t}) + E(\lambda_{i,t}) \\ = \lambda_{i,t}d_1 \left(\frac{1}{(\mu - \lambda_{i,t})} - d_0 \right)^+ + e_1\lambda_{i,t} + e_0 \quad (1)$$

可见操作能耗函数 $f(\lambda)$ 是非负且非递减的凸函数。

其次推导切换能耗函数，由于活跃服务器切换到休眠模式需要负载迁移、机器折旧等损耗，而休眠模式到活跃模式的能耗极小可以忽略。切换能耗发生在相邻时隙 $t-1$ 和 t 之间，表达为 $\beta(x_t - x_{t-1})^+$ ，其中切换系数 β 为正常数。由于“负载均衡”调度策略被广泛接收是最优分配方式^[6,12]，则任意服务器 $i \in [1, 2, \dots, x_t]$ 被分配到的负载为 $\lambda_{i,t} = \lambda_t/x_t \leq M$ ，且有 $x_0 = x_T = 0$ 。综上所述，数据中心能耗最小化问题可表述为

问题 1

$$\left. \begin{aligned} \min_{x_1, x_2, \dots, x_{T-1}} \quad & \sum_{t=1}^{T-1} x_t f(\lambda_t/x_t) + \beta \sum_{t=1}^{T-1} (x_t - x_{t-1})^+ \\ \text{s.t.} \quad & x_t \in \lceil \lambda_t/M, N \rceil \end{aligned} \right\} \quad (2)$$

其中 x_t 是待优化整数变量， T, β, N, M 是常量， $f(\cdot)$ 是凸函数。离线方式下， λ_t 是统计获取的常量；在线方式， λ_t 是预测的变量。显然问题1为整数规划问题，而此种整数规划问题是典型NP-hard问题^[17]。

3 最优解特性分析

3.1 无切换成本最优解

式(2)的解可表述为整数序列 $\varphi^* = (x_1^*, x_2^*, \dots, x_{T-1}^*)$ ，首先考虑特殊情况 $\beta = 0$ (无切换成本)。记

$$F(x) = xf(1/x), \quad x \geq 1 \quad (3)$$

由于 $f(x)$ 是凸函数， $F(x)$ 也是凸函数。因而操作能耗可写为 $\sum_{t=1}^{T-1} x_t f(\lambda_t/x_t) = \sum_{t=1}^{T-1} \lambda_t (x_t/\lambda_t) f(\lambda_t/x_t)$

$= \sum_{t=1}^{T-1} \lambda_t F(x_t/\lambda_t)$ ，进而问题 1 简化为问题 2。

问题 2

$$\min_{x_1, x_2, \dots, x_{T-1}} \sum_{t=1}^{T-1} \lambda_t F(x_t/\lambda_t), \quad \text{s.t. } x_t \in [\lambda_t/M, N] \quad (4)$$

x_t 为整数变量， T, λ_t, N, M 为常量。

问题 2 可拆分成 $T-1$ 个关于 x_t 的子问题，即最小化 $\lambda_t F(x_t/\lambda_t)$ ， $1 \leq t \leq T-1$ 。由于 $F(x)$ 是凸函数，定义真值 x^F 使 $x \in [1, NM/\lambda_t]$ 区间内 $F(x)$ 最小。从而得到最优解 $x_t^* \approx x^F \lambda_t/M$ 在约束条件下 x_t^* 是绑定于 $[\lambda_t/M]$ 和 N 的整数。此时最优解 $x_t^* = x_t^F$ ，且 x_t^F 表达为：

$$x_t^F = \begin{cases} \lfloor \lambda_t \rfloor, & x^F \lambda_t \leq \lfloor \lambda_t \rfloor \\ \lfloor x^F \lambda_t/M \rfloor, & \lfloor \lambda_t/M \rfloor < x^F \lambda_t/M < N, \\ & F(\lfloor x^F \lambda_t/M \rfloor) \leq F(\lfloor x^F \lambda_t/M \rfloor + 1) \\ \lfloor x^F \lambda_t/M \rfloor, & \lfloor \lambda_t/M \rfloor < x^F \lambda_t/M < N, \\ & F(\lfloor x^F \lambda_t/M \rfloor) \leq F(\lfloor x^F \lambda_t/M \rfloor + 1) \\ N, & x^F \lambda_t = NM \end{cases} \quad (5)$$

3.2 一般最优解特性

当一般情况下 $\beta > 0$ 时，初始 $x_0^* = 0$ ，为了获得问题 1 的最优解，直观的思路是小心地选择每个 x_t^* ，使其与 x_t^F, x_{t+1}^F 偏差引起的切换能耗和操作能耗均最小，给出引理 1-引理 3。鉴于篇幅，证明过程省略。

引理 1 如果 $x_t^* = x_t^F$ 在最优解 φ^* 中，则 x_{t+1}^* 是处在 x_t^* 与 x_{t+1}^F 之间的值。

引理 2 如果 $x_t^* \neq x_t^F$ 在最优解 φ^* 中，且 x_{t+1}^F 和 x_t^F 在 x_t^* 同一侧 (x_{t+1}^F 和 x_t^F 同时小于等于 x_t^* 或同时大于等于 x_t^*)，则 $x_{t+1}^* = x_t^*$ 。

引理 3 如果 $x_t^* \neq x_t^F$ 在最优解 φ^* 中，且 x_{t+1}^F 和 x_t^F 不在同一侧 (x_{t+1}^F 和 x_t^F 不同时小于 x_t^* 或不同时不大于 x_t^*)，则 x_{t+1}^* 是处在 x_t^* 和 x_{t+1}^F 之间的值。

根据以上 3 条引理，可得到从 $t = T$ 到 $t = 0$ 时间段内最优曲线 (t, x_t^*) 与 (t, x_t^F) 曲线的关系。得到一个跟随周期 $t \in [\tau, T]$ ，因为在 $t = T$ 时刻 $x_t^* = x_t^F$

$= 0$ 。一旦得到 $x_{\tau-1}^* \neq x_{\tau-1}^F$ 则 $x_{\tau-1}^* = x_{\tau-2}^* = \dots = x_{\tau}^*$ ，直到这条平整线与 (t, x_t^F) 曲线相交，称 $[\hat{\tau}, \tau - 1]$ 为平周期。进而将得到下一个跟随周期和平周期直至结束。跟随周期和平周期如图 1 所示。

4 离线最优算法

给定整数 $w \in [0, T]$ 和 $h \in [\lambda_t/M, N]$ ，定义在指定结束时隙 w 和结束时隙内活跃服务器数为 h ，最小化整体能耗的一组子问题。表述如问题 3。

问题 3

$$\min_{x_0, x_1, \dots, x_w} \left. \begin{aligned} \text{CT}(w, h) &= \sum_{t=0}^w x_t f(\lambda_t/x_t) + \beta \sum_{t=1}^w (x_t - x_{t-1})^+ \\ \text{s.t. } x_w &= h, \quad x_t \in [\lambda_t/M, N], \quad t = 0, 1, \dots, w \end{aligned} \right\} \quad (6)$$

令问题 3 达到最优解时的整体能耗为 $\text{CT}^*(w, h)$ ，对于目标问题 1，已知有 $x_T = 0$ ，因此目标问题 1 的整体最小能耗定义为 $\text{CT}^*(T, 0)$ 。引理 1~引理 3 同样适用于问题 3。定义 $\text{CT}_{(w,h),(v,j)}$ 为时隙序列 (w, h) 到 (v, j) 能耗之和。动态规划过程分为以下 3 个连续承接的步骤：

第 1 步 $\text{CT}^*(T, 0)$ 递推：根据式(5)和 $\lambda_T = 0$ ，有 $x_T^F = 0$ ，于是总能找到整数

$$t_1 = \min \{s \mid 1 \leq s \leq T, x_s^F \geq x_{s+1}^F \geq \dots \geq x_T^F\} \quad (7)$$

如图 2 所示，绘出一段 (t, x_t^F) 的曲线，在 $[t_1, T]$ 段 $x_{t_1}^F \geq x_{t_1+1}^F \geq \dots \geq x_T^F = 0$ 。由于 $x_T^* = 0 = x_T^F$ ，在一个最优解 φ^* 中，存在一个 $\tau_1 \in [t_1 + 1, T]$ 和 $h_1 \in [1, x_{\tau_1-1}^F - 1]$ ，使在 $t \in [\tau_1, T]$ 内， $x_t^* = x_t^F$ 且 $x_{\tau_1-1}^* = h_1 < x_{\tau_1-1}^F$ ，即出现 3.2 节定义的最后平周期和最后一个跟随周期。与此同时，定义 $\hat{\tau}_1$ 是 $[1, t_1]$ 上，高度为 h_1 的平等线与曲线 (t, x_t^F) 的交点。即

$$\hat{\tau}_1 = \max \{s \mid 1 \leq s \leq t_1, x_s^F = h_1 \neq x_{s-1}^F\} \quad (8)$$

设 $x_{\hat{\tau}_1}$ 的值固定为 h_1 时在区间 $t \in [0, \hat{\tau}_1]$ 上的整体能耗最小值为 $\text{CT}^*(\hat{\tau}_1, h_1)$ 。由于在 $t \in [\hat{\tau}_1, \tau_1 - 1]$ 上 $x_t^* = h_1$ 且 $t \in [\tau_1, T]$ 上 $x_t^* = x_t^F$ 。可计算出 $t \in [\hat{\tau}_1, T]$ 上的整体总能耗(不包括 τ_1 时隙的操作能耗，因为此能耗已在 $\text{CT}^*(\hat{\tau}_1, h_1)$ 中被计算过) $\text{CT}_{(\hat{\tau}_1, h_1), (T, 0)}$ 。因而

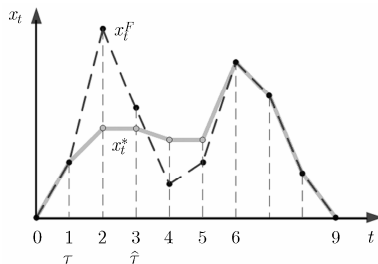


图 1 最优解特性—跟随周期与平周期

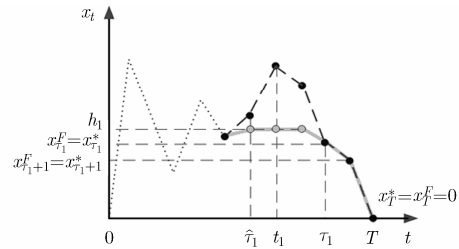


图 2 递推用例

得到 $CT^*(T, 0) = CT^*(\hat{\tau}_1, h_1) + CT_{(\hat{\tau}_1, h_1), (T, 0)}$, 扩展到一般情况时, 得到

$$CT^*(T, 0) = \min_{h_1 \in [1, x_1^F]} (CT^*(\hat{\tau}_1, h_1) + CT_{(\hat{\tau}_1, h_1), (T, 0)}) \quad (9)$$

显然, 式(9)给出了一个基于一些 $CT^*(\hat{\tau}_1, h_1)$ 值的计算 $CT^*(T, 0)$ 递推方法。

第2步 $CT^*(\hat{\tau}_1, h_1)$ 计算: 定义

$$t_2 = \min \{s \mid 1 \leq s \leq \hat{\tau}_1, x_s^F \leq x_{s+1}^F \leq \dots \leq x_{\hat{\tau}_1}^F\} \quad (10)$$

在 $CT^*(\hat{\tau}_1, h_1)$ 的最优解 x_t^* 中必会出现以下两种

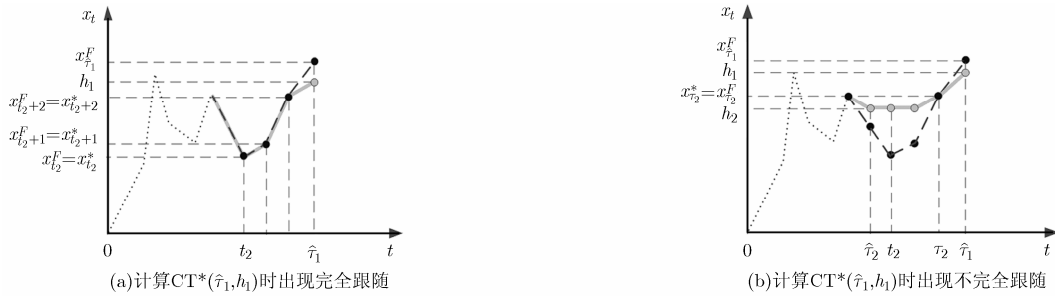


图3 计算 $CT^*(\hat{\tau}_1, h_1)$ 时存在的两种情况

$$\hat{\tau}_2 = \max \left\{ s \mid 1 \leq s \leq t_2, x_{\hat{\tau}_2}^F = h_2 \neq x_{\hat{\tau}_2-1}^F \right. \\ \left. \text{or } x_{\hat{\tau}_2}^F < h_2 < x_{\hat{\tau}_2-1}^F \right\} \quad (11)$$

$x_{\hat{\tau}_2}^*$ 固定为 h_2 时最小总能耗为 $CT^*(\hat{\tau}_2, h_2)$ 。由于 $t \in [\hat{\tau}_2, \tau_2 - 1]$ 上 $x_t^* = h_2$, 在 $t \in [\hat{\tau}_2, \tau_1 - 1]$ 上 $x_t^* = x_t^F$, 且 $x_{\hat{\tau}_1}^* = h_1$, 可以计算出 $t \in [\hat{\tau}_2, \hat{\tau}_1]$ 上的总能耗 (不包括 $\hat{\tau}_2$ 时刻的操作能耗, 其已在 $CT^*(\hat{\tau}_2, h_2)$ 中被计算过) $CT_{(\hat{\tau}_2, h_2), (\hat{\tau}_1, h_1)}$, 于是有 $CT^*(\hat{\tau}_1, h_1) = CT^*(\hat{\tau}_2, h_2) + CT_{(\hat{\tau}_2, h_2), (\hat{\tau}_1, h_1)}$ 。

综合以上两种情况得

$$CT^*(\hat{\tau}_1, h_1) = \min_{h_2 \in [x_{t_2}^F, \min\{h_1, \max_{t=1}^{t_2} x_t^F\}]} \cdot (CT^*(\hat{\tau}_2, h_2) + CT_{(\hat{\tau}_2, h_2), (\hat{\tau}_1, h_1)}) \quad (12)$$

第3步 其他递推: 一旦一些 $CT^*(\hat{\tau}_2, h_2)$ 的值被确定, 式(12)中给出了计算 $CT^*(\hat{\tau}_1, h_1)$ 的递推公式, 可以继续递推过程。当确定在 $x_{\hat{\tau}_k-1}^F < h_k \leq x_{\hat{\tau}_k}^F$ 时 $CT^*(\hat{\tau}_k, h_k)$ 的值时, 过程终止, 并且此时 $x_0^F \leq x_1^F \leq \dots \leq x_{\hat{\tau}_k}^F$ 。对于这种情况, 很容易证明在 $t \in [1, \hat{\tau}_k - 1]$ 上满足 $CT^*(\hat{\tau}_k, h_k)$ 的最优解是 $x_t^* = x_t^F$, 且 $x_{\hat{\tau}_k}^* = h_k$ 由此终止了递推过程。如表1所示。

5 在线优化算法

在线算法的求解目标是根据 x_{t-1} 和 x_t^F 的值来确定 x_t 值, x_{t-1} 的值根据之前的 λ_τ 值确定 ($\tau < t$), x_t^F 根据当前 λ_t 的值由式(5)计算得到。为此给出引理4。

引理4 给定 x_{t-1} 和 λ_t , 最优在线算法应将 x_t 设置为 x_{t-1} 和 x_t^F 中的某个值, 其中 x_t^F 由式(5)确定。

情况中的一种: (1)对于所有的 $t \in [t_2, \hat{\tau}_1 - 1]$, $x_t^* = x_t^F$ (见图3(a)); (2)存在一个 $\tau_2 \in [t_2 + 1, \hat{\tau}_1 - 1]$ 和 $h_2 \in [x_{t_2}^F + 1, h_1]$ 使得对于 $t \in [\tau_2, \hat{\tau}_1 - 1]$ 有 $x_t^* = x_t^F$, 且 $x_{t_2-1}^* = h_2 > x_{t_2-1}^F$ (见图3(b))。对于情况(1), 显然: $CT^*(\hat{\tau}_1, h_1) = CT^*(t_2, x_{t_2}^F) + CT_{(t_2, x_{t_2}^F), (\hat{\tau}_1, h_1)}$ 。对于情况2, 可以证明对于任意的 $h_2 > \max_{t=1}^{t_2} x_t^F$ 都不能产生出最优解。由于 $h_2 \in [x_{t_2}^F + 1, \min\{h_1, \max_{t=1}^{t_2} x_t^F\}]$, 定义 $\hat{\tau}_2$ 是高度为 h_2 的平高线与曲线 (t, x_t^F) 的交点, 亦即

(1)当 $x_{t-1} = x_t^F$ 时, 根据引理4得 $x_t = x_{t-1}$ 。

(2)当 $x_{t-1} > x_t^F$ 时, 根据引理4得到 x_t 应在 $[x_t^F, x_{t-1}]$ 之间。由于之后的 λ_τ ($\tau \geq t + 1$) 的值未知, 无法根据未知的 λ_τ 的值来确定具有最优性能的 x_t 的值。此处考虑 λ_{t+1} 为最坏的情况下时求解具有最优性能的 x_t 。

$t \leq T - 2$ 时, 最坏的情况是 $\lambda_{t+1} \geq x_{t-1}$, 其中 $x_{t+1} \geq \lambda_{t+1} \geq x_{t-1} > x_t^F$ 。需要在区间 $[x_t^F, x_{t-1}]$ 里选择一个合适的 x_t 的值。通过将 x_t 的值从 h 增加到 $h + 1$, 其中 $h \in [x_t^F, x_{t-1} - 1]$, 在区间 $[t - 1, t + 1]$ 内

表1 离线最优算法伪代码

算法1 基于动态规划的数据中心能耗最优算法

- (1)根据式(5)计算 x_t^F , $t \in [1, T - 1]$
- (2)找到 x_t^F 的第一个和最后一个局部最大点对应的 U_1 和 U_K 。
- (3) for ($t = 1; t \leq U_1; t++$)
- (4) if $x_t^F > x_{t-1}^F$
- (5) 计算 $CT(t, h)$, $h \in [x_{t-1}^F + 1, x_t^F]$, 其中 $x_\tau^* = x_\tau^F$, $\tau \in [1, t - 1]$, 且 $x_t^* = h$
- (6) for ($t = U_1 + 1; t \leq U_K; t++$)
- (7) if $x_t^F < x_{t-1}^F$
- (8) 计算 $CT(t, h)$, $h \in [x_t^F, x_{t-1}^F - 1]$, 通过之前给定的 $CT(\cdot, \cdot)$, 参照式(9)。
- (9) else if $x_t^F < x_{t-1}^F$
- (10) 计算 $CT(t, h)$, $h \in [x_{t-1}^F + 1, x_t^F]$, 通过之前给定的 $CT(\cdot, \cdot)$, 参照式(12)。
- (11) 计算 $CT(T, 0)$ 通过之前给定的 $CT(\cdot, \cdot)$, 参照式(9)。

的切换能耗总是减少的。 $f(\cdot)$ 是凸函数, t 时隙时的操作能耗, 即 $\lambda_t \left[f\left(\frac{h+1}{x_{\lambda,t}}\right) - f\left(\frac{h}{x_{\lambda,t}}\right) \right]$ 会随着 h 的增加而增加。因此, x_t 从 x_t^F 到 x_{t-1} 时增加的总能耗为

$$m(h) = \beta - \lambda_t \left[F\left(\frac{h+1}{x_{\lambda,t}}\right) - F\left(\frac{h}{x_{\lambda,t}}\right) \right] \quad (13)$$

在开始($h = x_t^F$)时, 该值可能为正值; 当到达某个 h^* 时, 数值转为负的。因此可以设置 $x_t = h^*$, 而在一些极限情况下, 可以取 $x_t = x_t^F$ (如果 $m(x_t^F) \leq 0$) 或 $x_t = x_{t-1}$ (如果 $m(x_{t-1}) > 0$)。

$t = T - 1$ 时, 则不存在未知的 λ_T 值。因为 $\lambda_{T+1} = \lambda_T = 0$, 所以 $x_{T+1} = 0$ 。由于 $x_{t-1} \geq x_t > x_{t+1} = 0$ 。可以令 t 时隙时 $x_t = x_t^F$ 来使切换能耗和操作能耗在区 $[t-1, t+1]$ 内同时达到最小。

(3) 当 $x_{t-1} < x_t^F$ 时, 最坏的情况为 $\lambda_{t+1} \ll x_{t-1}$ 。此时 $x_{t+1}^F \ll x_{t-1}$ 且 x_{t+1} 不大于 x_{t-1} 。一方面, 可以在时隙 t 时设置 $x_t = x_t^F$ 使操作能耗最小; 另一方面, $x_t^F > x_{t-1} \geq x_{t+1}$, 所以 x_t 越接近 x_{t-1} 切换能耗越小。类似于情况(2)中的第 1 种情况, 能够在区间 $[x_{t-1}, x_t^F]$ 内为 x_t 取得合适的值。整体算法流程如表 2 所示。

表 2 在线优化算法伪代码

算法 2 基于最坏预测情况的在线算法

```

(1) for ( $t = 1; t < T; t++$ ) {
(2)   if ( $x_{t-1} == x_t^F$ ) 令  $x_t = x_{t-1}$ 
(3)   else if ( $x_{t-1} > x_t^F$ ) {
(4)     if ( $t \leq T - 2$ ) {
(5)       if ( $m(x_t^F) \leq 0$ ) 令  $x_t = x_t^F$ 
(6)       else if ( $g(x_{t-1}) \geq 0$ ) 令  $x_t = x_{t-1}$ 
(7)       else 确定一个  $h^* \in [x_t^F + 1, x_{t-1} - 1]$  使  $m(h^*) \geq 0$  且  $m(h^* + 1) < 0$ 
(8)         令  $x_t = h^*$ 
(9)     } else 令  $x_t = x_t^F$ 
(10)   } else {
(11)     if ( $m(x_{t-1}) \leq 0$ ) 令  $x_t = x_{t-1}$ 
(12)     else if ( $m(x_t^F) \geq 0$ ) 令  $x_t = x_t^F$ 
(13)     else 确定一个  $h^* \in [x_{t-1} + 1, x_t^F - 1]$  使得  $m(h^*) \geq 0$  且  $m(h^* + 1) < 0$ 
(14)       令  $x_t = h^*$ 
}

```

在线算法的每一步需要求解当前的 x_t^F , 其复杂度为 $O(N)$ 。然后需要根据前一时间片的取值情况确定本次情况。当 $x_{t-1} == x_t^F$ 时直接得到 $x_t = x_{t-1}$, 复杂度为 $O(1)$; 当 $x_{t-1} > x_t^F$ 时再根据 $m(x_t^F)$, $g(x_{t-1})$ 作出相应判断, 其复杂度为 $O(1 + 1 + 1) = O(1)$; 对于最后一种情况 $h^* \in [x_{t-1} + 1, x_t^F - 1]$, 此时可在区间内任取一点, 复杂度仍然为 $O(1)$ 。综上, 在线算法的复杂度为 $O(N)$ 。

6 仿真数值结果

由于条件所限, 仿真实验在 matlab 2013a 环境下, 采用离散事件动态方法进行仿真。整体运行模式与流程类似于数据中心的模型设定: 服务器数 $N = 100$, 服务容量 $M = 1$, 时隙总长 $T = 100$; 能耗参数分别设为 $d_1 = 1, \mu = 1, e_1 = 0, e_0 = 1$, 则操作能耗函数为 $f(\lambda) = \lambda/(1 - \lambda) - 1.5, 0)^+ + 1$ 。

6.1 离线仿真分析

离线情况下时各时隙 $t \in [1, 2, \dots, 99]$ 内的负载 λ_t 随机取 $[1, N] = [1, 100]$ 上的整数。选定 $\beta = 6$ 时, 比较两种负载到达模式时的效果。如图 4 所示, 对于各时隙平稳波动负载和剧烈波动负载, (t, x_t^F) 曲线明显和 (t, λ_t) 同步, 尤其是在剧烈波动情况下系统服务器使用率接近 90% 而且频繁切换; 如图 5 所示, 将每 5 个时隙合取成 X 坐标, 采用最优算法之后, 即使在剧烈波动情况下, x_t^{*2} 结果仍然维持且在各阶段的总能耗比 x_t^{F2} 结果大约降低 30%。

6.2 在线仿真分析

在线情况下, 各时隙 $t \in [1, 2, \dots, 99]$ 内的仿真负载 $\tilde{\lambda}_t = k(\lambda_t)^\alpha$, 通过参数 k 和 α 加以控制, 此处取 $k \in [1, 2]$ 和 $\alpha \in [2, 4]$ 用于随机生成各时隙的动态用户到达数。首先根据式(5)计算出 x_t^F , 将离线算法和在线算应用于此场景。图 6 显示离线算法和在线算法同时运行所得到的曲线, 可见活跃服务器总是可以在每个时隙 t 完成相应的工作负载(即 $x_t^*, x_t \geq \lambda_t$)。定义在线算法的总能耗与离线算法的总能耗的比值为“性能比”, 该比值总是不小于 1 的值, 当在线算法的性能越逼近近似最优解, 则性能比应越接近于 1。图 6 中离线算法所用总能耗为 14976.8, 在线算法所用总能耗为 17748.5, 在线算法的性能比为 $\frac{17748.5}{14976.8} = 1.185$, 较为接近 1, 即局部最优较为近似最优解。文献[16]采用的 Lazy 算法, 能够保证各周期内的活跃服务器数量与本文在线算法效果接近, 然而从图 6 中看出其最优值不能保证服务器为整数。

图 7 中 online 所示为 100 种场景下所得到的在线算法“性能比”曲线, 可见其非常接近 1。100 种场景下的平均“性能比”为 1.151, 其中最大值和最小值分别为 1.165 和 1.133。将 x_t^F 作为一种在线算法进行比较, 所得到的“性能比”并不接近 1, 因为其仅仅最小化了操作能耗。 x_t^F 的“性能比”曲线其平均性能比为 2.197, “性能比”最大值和最小值分别为 2.297 和 2.111。而 Lazy 算法的“性能比”始终稳定在 1.45, 比本文的在线优化算法高出 10%。由此可见, 同时考虑操作能耗和切换能耗是十分必

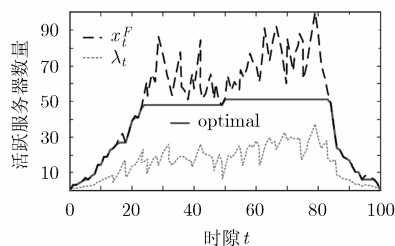


图4 不同负载变化情况下活跃服务器数

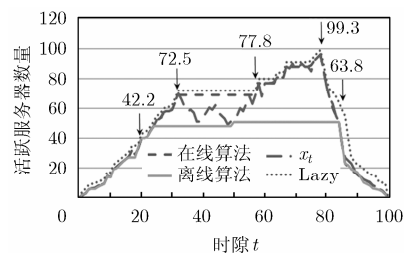


图6 在线算法与最优离线算法的比较

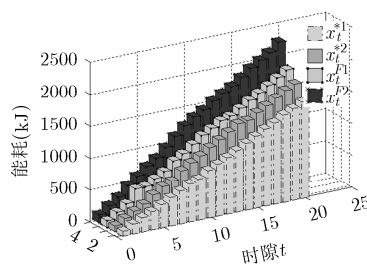


图5 不同负载变化情况下系统最小能耗

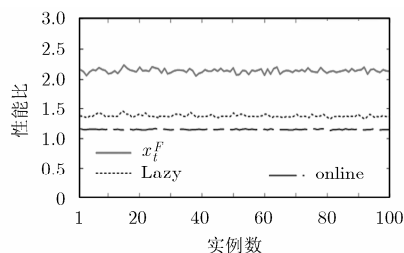


图7 100组工作负载场景下“性能比”曲线

要的，两者必须同时达到均衡点才能使总体能耗最接近最优离线算法所得到的最优解。

7 结束语

本文研究如何静态(离线)/动态(在线)配置连续运行时隙的活跃服务器数量，以最小化数据中心能耗的问题。数值结果表明，本文所提出的离线最优算法以较低的复杂度缩短了连续时隙运行时延，同时符合活跃服务器数量需为整数的要求，为在线算法提供最优参考依据。仿真分析表明，本文提出的在线优化算法，能够动态适应外界负载的剧烈变化趋势，始终较为谨慎地调整活跃服务器和休眠服务器的比例，始终以接近最优的能耗代价维持数据中心的平稳运行。进一步的工作可以分为两方面，一是以实际云计算数据中心的真实海量数据为来源，印证和提高算法的可行性与实用性，二是研究负载调度与活跃服务器配置联合的综合策略。

参考文献

- [1] Chong F T, Heck M J R, Ranganathan P, *et al.*. Data center energy efficiency: improving energy efficiency in data centers beyond technology scaling[J]. *IEEE Design & Test*, 2014, 31(1): 93-104.
- [2] Li Jian, Shuang Kai, Su Sen, *et al.*. Reducing operational costs through consolidation with resource prediction in the cloud[C]. 12th IEEE/ACM International Symposium on Cloud and Grid Computing (CCGrid), Ottawa, Canada, 2012: 793-798.
- [3] Wang Lin, Zhang Fa, Arjona Aroca J, *et al.*. GreenDCN: a general framework for achieving energy efficiency in data center networks[J]. *IEEE Journal on Selected Areas in*

Communications, 2014, 32(1): 4-15.

- [4] Uргаonkar R, Kozat U C, Igarashi K, *et al.*. Dynamic resource allocation and power management in virtualized data centers[C]. *IEEE/IFIP Network Operations and Management Symposium (NOMS)*, Osaka, Japan, 2010: 479-486.
- [5] Guenter B, Jain N, and Williams C. Managing cost, performance, and reliability tradeoffs for energy-aware server provisioning[C]. 2011 Proceedings of IEEE International Conference on Computer Communications (INFOCOM), Shanghai, China, 2011: 1332-1340.
- [6] Qureshi A, Weber R, Balakrishnan H, *et al.*. Cutting the electric bill for internet-scale systems[J]. *ACM SIGCOMM Computer Communication Review*, 2009, 39(4): 123-134.
- [7] Guo Yuan-xiong and Fang Yu-guang. Electricity cost saving strategy in data centers by using energy storage[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2013, 24(6): 1149-1160.
- [8] Rao Lei, Liu Xue, Xie Le, *et al.*. Minimizing electricity cost: Optimization of distributed internet data centers in a multi-electricity market environment[C]. 2010 Proceedings of IEEE International Conference on Computer Communications (INFOCOM), San Diego, CA, USA, 2010: 1-9.
- [9] Cao Jun-wei, Li Ke-qin and Stojmenovic I. Optimal power allocation and load distribution for multiple heterogeneous multi-core server processors across clouds and data centers[J]. *IEEE Transactions on Computers*, 2014, 63(1): 45-58.
- [10] Beloglazov A, Buyya R, Lee Y C, *et al.*. A taxonomy and survey of energy-efficient data centers and cloud computing systems[J]. *Advances in Computers*, 2011, 82(2): 47-111.
- [11] Wang Kai, Lin Ming-hong, Ciucu F, *et al.*. Characterizing the impact of the workload on the value of dynamic resizing in data centers[C]. *ACM SIGMETRICS/Performance*, London,

- United Kingdom, 2012: 405-406.
- [12] Rabbani M G, Zhani M F, and Boutaba R. On achieving high survivability in virtualized data centers[J]. *IEICE Transactions on Communications*, 2014, E97B(1): 10-18.
- [13] Liu Zhen-hua, Lin Ming-hong, Adam W, *et al.* Greening geographical load balancing[C]. Proceedings ACM SIGMETRICS, San Jose, CA, USA, 2011: 233-244.
- [14] Mathew V, Sitaraman R K, and Shenoy P. Energy-aware load balancing in content delivery networks[C]. Proceedings of the ACM SIGMETRICS Joint International Conference on Measurement and Modeling of Computer Systems, Orlando, FL, USA, 2012: 954-962.
- [15] Gandhi A, Gupta V, Harchol Balter M, *et al.* Optimality analysis of energy-performance trade-off for server farm management[J]. *Performance Evaluation*, 2010, 67(11): 1155-1171.
- [16] Lin Ming-hong, Wierman A, Andrew L L H, *et al.* Dynamic right-sizing for power-proportional data centers[J]. *IEEE/ACM Transactions on Networking*, 2013, 21(5): 1378-1391.
- [17] Michael R G and Johnson D S. Computers and Intractability: A Guide to the Theory of NP-completeness[M]. San Francisco: WH Freeman & Co., 1979: 206-218.
- 卫 星: 男, 1980年生, 博士后, 主要研究方向为计算机网络、离散事件动态性能优化.
- 张建军: 男, 1963年生, 教授, 主要研究方向为机电一体化、物联网工程、新能源汽车、汽车电子.
- 石 雷: 男, 1980年生, 讲师, 主要研究方向为无线传感网.
- 翟 琰: 女, 1977年生, 讲师, 主要研究方向为汽车电子、嵌入式系统.