

电力物联网云主站计算负荷模型与资源优化配置

孙宇嫣, 蔡泽祥, 马国龙, 郭采珊

(华南理工大学 电力学院, 广东 广州 510640)

摘要:随着电力物联网建设不断发展,运行分析、电能交易、用户能效管理、电源管理等业务应用呈现指数级增长,给传统主站系统带来巨大的计算处理压力。为了解决上述问题,针对新一代“云主站”架构,建立了电力物联网的计算负荷模型,并基于此提出了云资源配置方法。从“云主站”内计算负荷动态变化的根源着手,利用马尔科夫链刻画电网运行状态的变化规律;根据业务处理流程规范,建立状态关联的计算负荷模型,包括计算负荷描述模型及状态-事件-应用关联模型,并将其建立为经验知识库,用于分析不同时段各类业务的计算负荷特征;基于上述计算负荷模型,以平均响应延时最短、“云主站”能耗最小为目标,建立云资源优化配置双目标优化模型,通过逼近理想解排序法从帕累托前沿解集中获得综合最优解。以改进的IEEE 33节点电网为例,仿真结果验证了所提方法能有效分析不同应用计算负荷的变化特征,进而合理配置云资源。

关键词:电力物联网;云主站;计算负荷模型;云资源配置;马尔科夫链;经验知识库;逼近理想解排序法

中图分类号:TM 71

文献标志码:A

DOI:10.16081/j.epae.202102027

0 引言

互联网技术和新能源技术的推动作用下,新一轮能源革命以智能电网与“互联网+”技术相结合的形式成为了当前研究领域的热点^[1-2]。“云主站”作为电力物联网(PIoT)信息汇总和应用处理的中枢,其接收的业务呈现海量信息接入、并发用户交互、实时闭环控制等复杂多样化的趋势,必定给主站内部的计算决策系统带来更加严峻的挑战。随着智能传感、物联网、5G通信、大数据等技术不断成熟并广泛应用,电网自动化架构将从传统的主子站模式逐渐发展成为基于云计算技术架构的新模式^[3-4]。

借鉴云计算的资源虚拟化技术,相关学者提出了“云主站”的概念,即在主站平台应用云技术部署多种微服务,打破传统垂直式架构^[5-6]。文献[7]将给定时段内到达数据中心服务器的工作总量定义为计算负荷,而业务应用可映射为“云主站”的计算负荷,每一时段对“云主站”内部资源配置提出了不同的要求。如何根据业务计算负荷需求有效分配云资源,以满足业务需求并确保“云主站”经济高效运行,成为学术界、工业界广泛关注的热点问题。在研究分析计算负荷需求方面,文献[8]明确了计算负荷需求分析的重要性,指出基于计算负荷特征的资源配置方法能使云中心掌握计算负荷走向,调整虚拟计算资源配置;文献[9-10]对面向移动用户的云中心记录的计算负荷数据,采用深度学习学习方法学习不同时间尺度下计算负荷数据的特征,并将其用于后续

的资源配置,但结果大多依赖于计算负荷单方面的显式特征,不能充分表达历史数据隐含的特征;文献[7,11-12]基于排队论、网络演算、动力学滤波等理论对计算负荷在CPU需求、I/O需求等方面的数据进行假设和模拟,进而模拟计算负荷模式以进行资源配置,但实际计算负荷特征时并不完全与假设模型相同,实际使用时计算负荷模型的可用性受到影响。

考虑到电力物联网业务与电网运行状态的强关联性及其业务计算负荷需求特征的差异性,仅靠计算负荷信息本身或直接套用理论模型无法准确表征不同业务的计算负荷需求变化情况。因此,本文提出电力物联网的计算负荷模型,并将其应用于“云主站”计算资源的优化配置。首先,通过马尔科夫链刻画电网系统运行状态的变化规律。然后,根据电力物联网业务处理流程规范,建立状态关联的计算负荷模型,一方面刻画不同类型业务应用在计算量、任务到达频率等方面的固有特征,另一方面利用状态-事件-应用的关联关系,融入基于概率的电网运行状态时间分布特征,修正得到计算负荷需求的时间分布特征。基于此,以平均响应延时、总能耗为目标建立云资源优化配置模型,利用CPLEX求解器和逼近理想解排序法(TOPSIS)得到综合最优解。算例结果表明:本文方法能合理配置云计算资源,减少“云主站”的响应时间、能耗。本文工作能为基于云计算架构的新型自动化模式的资源优化配置提供定量分析模型和方法,将有助于这一新模式的成熟和发展。

1 “云主站”计算资源优化配置的基本思路

针对上述问题及背景,本文提出了电力物联网“云主站”计算负荷模型及云资源优化配置方法,基本思路见图1。本文主要利用的数据为电力系统对

收稿日期:2020-05-18;修回日期:2020-12-27

基金项目:广东省重点领域研发计划项目(2019B111109002)
Project supported by the Key-area Research and Development Program of Guangdong Province(2019B111109002)

象相关数据,包括电网区域单元的运行信息、业务描述信息,前者包括电气量信息、环境量信息、设备健康状态信息等,用于刻画电网运行状态的变化规律;后者包括业务相关应用程序的功能配置信息、任务计算量、任务平均到达频率等信息,用于建立面向电力物联网业务应用的计算负荷描述模型。经验知识库(KBS)根据业务处理流程规范,将电网内状态-事件-应用间的关联关系表达为规则,建立状态-事件-应用关联模型,用于进一步表征计算负荷的事件分布特征,然后将结果输送至“云主站”代理(Broker)中。

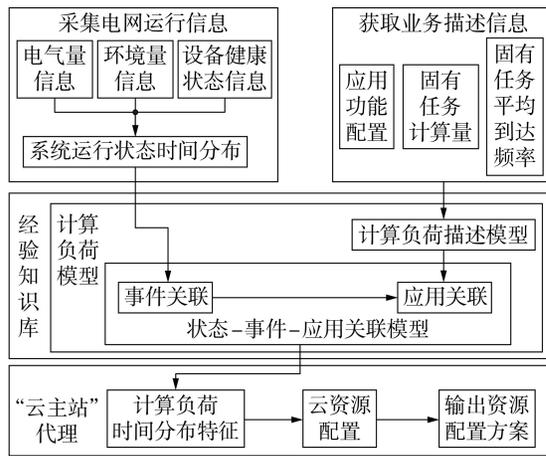


图1 计算负荷模型与云资源优化配置方法的基本思路

Fig.1 Basic idea of workload model and cloud resource's optimal allocation method

Broker的主要功能是感知云内工作环境的变化,实现对云内计算资源的有效管理与分配。在接收到计算负荷分布特征后,Broker在不同类型业务、不同的延时约束下,通过优化配置面向不同业务服务的虚拟机VM(Virtual Machine)计算资源,综合降低“云主站”内部的平均响应延时和总能耗。

2 电力物联网的计算负荷模型

为了保证电网高效运行及满足分析控制、区域内电力用户信息交互及多主体电力交易的需求,需要针对电网运行状态的评估情况,根据规范的业务处理流程确定该状态下对应的各类业务的计算需求,保证云中心资源能够合理地分配计算资源。

2.1 电网运行状态模型

2.1.1 数据表达

电网在运行过程中产生多时间尺度和多空间尺度的数据,这些数据隐含了运行状态的变化规律。为了深度挖掘运行状态的转换模式,需要对运行数据进行统一化表达。在运行的第 j 天 k 时刻的时间断面下,历史运行数据集 $Y_{k,j}$ 为:

$$Y_{k,j} = \{E_{k,j}, S_{k,j}, C_{k,j}\} \quad (1)$$

其中, $Y_{k,j} \subset Y_j, Y_j = \{Y_{0,j}, Y_{1,j}, \dots, Y_{m,j}\}$ 为第 j 天电网运行数据集, m 为一天内的总采样点数; $E_{k,j}, S_{k,j}, C_{k,j}$

分别为第 j 天 k 时刻的电气量信息集合、环境量信息集合(包括温度、湿度等)、设备健康状态信息集合。

2.1.2 基于马尔科夫链的电网运行状态

在离散状态空间 $S = \{s_1, s_2, \dots\}$ 中,离散时间参数的随机过程为 $\{X_t | X_t \in S, t \geq 0\}$ 。给定采样时段为 $[0, T]$,采集时间点为 $0 = t_0 < t_1 < t_2 < \dots < t_k < \dots < t_m = T$,时段间隔 $\Delta t = t_{k+1} - t_k$ 。 t_k 时刻的状态仅由 t_{k-1} 时刻的状态决定,如式(2)所示。该随机过程具备“无记忆性”的特性,称为马尔科夫链。

$$p(X_{k+1} = s(t_{k+1}) | X_1 = s(t_1), X_2 = s(t_2), \dots, X_k = s(t_k)) = p(X_{k+1} = s(t_{k+1}) | X_k = s(t_k)) \quad (2)$$

其中, $p(\cdot)$ 为条件概率; X_k 为 t_k 时刻的状态变量; $s(t_k)$ 为 X_k 的具体状态, $s(t_k) \in S$,并在 Δt 内持续该状态。

对于由 n 个状态组成的状态空间 S 而言,其在 t_k 时刻的状态转移概率矩阵 P^{t_k} 如式(3)所示。

$$P^{t_k} = \begin{bmatrix} p_{11}^{t_k} & p_{12}^{t_k} & \dots & p_{1n}^{t_k} \\ p_{21}^{t_k} & p_{22}^{t_k} & \dots & p_{2n}^{t_k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1}^{t_k} & p_{n2}^{t_k} & \dots & p_{nn}^{t_k} \end{bmatrix} \quad (3)$$

显然,式(3)具有如下性质:

$$\begin{cases} p_{ij}^{t_k} \geq 0 & i, j = 1, 2, \dots, n \\ \sum_{j=1}^n p_{ij}^{t_k} = 1 & i = 1, 2, \dots, n \end{cases} \quad (4)$$

其中, $p_{ij}^{t_k}$ 为矩阵 P^{t_k} 的元素,表示 t_{k-1} 时刻的状态 s_i 转换至 t_k 时刻的状态 s_j 的转移概率。

基于马尔科夫链的运行状态模型包含以下元素。

(1)状态集合 S :集合了马尔科夫链模型中所有的状态。

(2)初始状态概率矩阵 π :表示马尔科夫链模型中状态变量在初始时刻 t_0 的概率矩阵。基于式(1)所示的电网历史数据,通过统计所有时间尺度和空间尺度的运行状态情况,可获得初始状态概率矩阵。

(3)状态转移概率矩阵集合 P :表示在采样时段内每一采样时间点的状态转移概率矩阵,即 $P = \{P^{t_1}, P^{t_2}, \dots, P^{t_m}\}$ 。

基于上述内容建立电网的运行状态模型,进一步计算每个时刻的状态概率分布。假设已知 t_{k-1} 时刻的状态概率分布 $Z(t_{k-1})$, t_k 时刻的状态转移概率矩阵 P^{t_k} ,可计算得到经过 Δt 后 t_k 时刻的状态概率分布 $Z(t_k)$ 为:

$$Z(t_k) = Z(t_{k-1}) P^{t_k} \quad (5)$$

$$Z(t_k) = \begin{bmatrix} z_1(t_k) \\ z_2(t_k) \\ \vdots \\ z_n(t_k) \end{bmatrix}^T = \begin{bmatrix} p(X_k = s_1) \\ p(X_k = s_2) \\ \vdots \\ p(X_k = s_n) \end{bmatrix}^T \quad (6)$$

且满足:

$$\sum_{i=1}^n z_i(t_k) = 1 \quad (7)$$

其中,当 $t_k=0$ 时, $\mathbf{Z}(t_k)$ 为初始状态概率矩阵 $\boldsymbol{\pi}$ 。

2.2 状态关联的计算负荷模型

不同的状态场景会引发不同的业务应用,且即使是同一种业务应用,其计算任务量、任务平均到达频率等也会各不相同。电力物联网的运行控制业务处理流程规范隐含了状态、事件、应用间的关联关系,可借助这些关系实现状态关联的计算负荷分析。

2.2.1 计算负荷描述模型

电网运行过程中存在监测、控制、分析、交易、管理等多类型业务应用,不同应用的业务执行逻辑各不相同,其计算量、任务到达频率也各不相同,见表1。

表1 典型的电力物联网业务应用

Table 1 Typical business applications in PIOT

类别	业务内容	业务逻辑
监测类	数据采集、数据预处理、状态估计、异常告警等	持续执行,不可中断
控制类	机组发电控制等	与控制对象的状态相关
分析类	拓扑分析、潮流计算、安全评估等	持续执行,不可中断
交易类	电力交易	与对象的交易行为相关
管理类	能量管理、储能策略等	与系统源荷平衡状态、线路状态等相关

为了刻画不同业务应用的计算负荷特征,本文建立了计算负荷描述模型 Q_{APP_r} ,如式(8)所示。

$$Q_{APP_r} = \{L_r, \boldsymbol{\lambda}_r\} \quad (8)$$

其中, $L_r = [L_r(t_0), L_r(t_1), \dots, L_r(t_m)]$ 为第 r 类业务应用的计算量分布, $L_r(t_k)$ 为 t_k 时刻第 r 类业务应用的计算量,其初始值为该业务应用的固有计算量 $I_{L_r}(t_k)$; $\boldsymbol{\lambda}_r = [\lambda_r(t_0), \lambda_r(t_1), \dots, \lambda_r(t_m)]$ 为第 r 类业务应用的任务到达频率分布, $\lambda_r(t_k)$ 为 t_k 时刻第 r 类业务应用的任务到达频率,其初始值为该业务应用的固有任务到达频率 $I_{\lambda_r}(t_k)$ 。

2.2.2 状态-事件-应用关联模型

将电力物联网的业务处理流程规范表达为知识,用于分析某一状态下的具体应用及其计算负荷,并可持续不断地扩展、更新、修改,形成完备、独立、可扩充的KBS^[13-14]。KBS内的知识以规则形式存在,表现形式为:If<condition>then<conclusion>。这一部分的规则主要分为事件关联、应用关联2类。

事件关联表示状态-事件关联关系,通过事件关联模型 F 表征,见式(9)。借助状态-事件关联关系,根据已知的 $\mathbf{Z}(t_k)$ 进一步分析各根源事件的概率分布。

$$\begin{cases} F(\mathbf{Z}(t_k)) = \{F_1(z_1(t_k)), F_2(z_2(t_k)), \dots, F_n(z_n(t_k))\} \\ F_i(z_i(t_k)) = \{f_{i,1}(z_i(t_k)), f_{i,2}(z_i(t_k)), \dots, f_{i,N_i}(z_i(t_k))\} \end{cases} \quad (9)$$

其中, $f_{j,i}(z_i(t_k))$ ($k=1, 2, \dots, m; i=1, 2, \dots, n; j=1, 2, \dots, N_i$)为 t_k 时刻状态 s_i 条件下发生第 j 个根源事件的概率, N_i 为状态 s_i 的根源事件总数量。经式(9)处理后输出不同状态的根源事件发生概率,见式(10)。

$$\begin{cases} A(t_k) = \{A_1(t_k), A_2(t_k), \dots, A_n(t_k)\} \\ A_i(t_k) = \{a_{i1}(t_k), a_{i2}(t_k), \dots, a_{iN_i}(t_k)\} \end{cases} \quad (10)$$

其中, $A_i(t_k)$ ($i=1, 2, \dots, n; k=1, 2, \dots, m$)为经处理后 t_k 时刻状态 s_i 所有根源事件的概率分布集合; $a_{ij}(t_k)$ ($j=1, 2, \dots, N_i$)为经处理后 t_k 时刻状态 s_i 下第 j 个根源事件的发生概率。

应用关联描述的是电网事件与业务应用间的逻辑关系用 W 表示,如式(11)所示。

$$W = \{w_1, w_2, \dots, w_N\} \quad (11)$$

其中, w_j ($j=1, 2, \dots, N$)为第 j 个根源事件的业务应用集合, $w_j = \{APP_{1j}, APP_{2j}, \dots, APP_{M_jj}\}$, APP_{qj} 为第 j 个根源事件的第 q 个业务应用, M_j 为第 j 个根源事件关联的业务应用总数, N 为根源事件总数。

基于事件-业务应用关联关系,根据已知的 t_k 时刻电网事件的发生概率集合 $A(t_k)$,进一步分析每类业务应用的触发概率,并将其概率修正至计算负荷参数中,以实现某时刻下考虑所有电网运行状态的业务应用计算负荷总量估计,修正公式为:

$$\hat{L}_r(t_k) = \sum_{i=1}^n \sum_{j=1}^{N_i} \gamma_{ijr} \eta_{jr} a_{ij}(t_k) I_{L_r}(t_k) \quad (12)$$

$$\hat{\lambda}_r(t_k) = \sum_{i=1}^n \sum_{j=1}^{N_i} \gamma_{ijr} \eta_{jr} a_{ij}(t_k) I_{\lambda_r}(t_k) \quad (13)$$

其中, $\hat{L}_r(t_k)$ 、 $\hat{\lambda}_r(t_k)$ 分别为 t_k 时刻第 r 类业务应用计算量、平均到达频率的修正值; η_{jr} 为第 j 个根源事件内第 r 类业务应用的关联系数,由式(11)决定,若 w_j 内含有第 r 类业务应用则 $\eta_{jr}=1$,否则 $\eta_{jr}=0$; γ_{ijr} 为状态 s_i 下第 j 个根源事件中第 r 类业务应用的0/1修正系数,其数值由KBS中状态-事件-应用关联模型确定,若状态 s_i 、第 j 个根源事件与第 r 类业务应用均相关联,则 $\gamma_{ijr}=1$,否则 $\gamma_{ijr}=0$ 。

3 云计算资源优化配置方法

3.1 云计算资源优化配置模型

业务应用到达“云主站”后,将其加载到VM中进行计算处理。文献[15-16]认为计算系统内的每台VM仅处理单一类型应用的计算负荷。考虑到能量管理业务应用的多样性,相应VM的计算能力也各不相同,且能适应计算负荷变化进行动态调整。

3.1.1 目标函数

“云主站”的运行性能主要考虑实时性和能耗。一方面,不同类型的电力物联网业务应用为电力系统提供全方位的服务,如网络运行评估、事故应急处理、装置动态控制以及市场服务等,这些业务处理的

实时性会影响整个系统的运行性能;另一方面,“云主站”处理业务应用时往往伴随着能耗,而能耗会影响“云主站”运行的经济性、环境友好程度及可靠性。因此,“云主站”的资源配置需要兼顾实时性和能耗两方面。则目标函数为:

$$\min G = \{g_1, g_2\} \quad (14)$$

其中, g_1 为“云主站”的实时性目标函数; g_2 为“云主站”的总能耗目标函数。

由于“云主站”为每类业务应用均配置1台VM,当VM内部处理来自电网不同位置的同类业务应用时,业务应用处理的实时性具体表现为计算时间和等待时间。根据排队论,每台VM内应用到达时间遵循泊松分布,且VM的服务时间遵循指数分布,因此每台VM可用M/M/1排队模型计算其平均响应延时,包括任务在VM中的等待时间以及VM处理任务的计算时间。因此,“云主站”的实时性目标函数主要由总平均响应延时构成,如式(15)所示。

$$g_1 = \sum_{r=1}^R \sum_{t=1}^T d_r(t) \quad (15)$$

其中, R 为业务应用的总数; $d_r(t)$ 为 t 时段内VM对第 r 类业务应用的平均响应延时。当计算负荷为0时,认为当前无第 r 类业务应用,平均响应延时为0;当 t 时段内存在第 r 类业务应用的计算负荷时,可根据M/M/1排队模型计算平均响应延时,计算公式如式(16)所示。

$$d_r(t) = \begin{cases} \frac{1}{u_r(t)/L_r(t) - \hat{\lambda}_r(t)} & \hat{\lambda}_r(t) \neq 0 \\ 0 & \hat{\lambda}_r(t) = 0 \end{cases} \quad (16)$$

其中, $u_r(t)$ 为 t 时段内“云主站”为第 r 类业务应用配置的VM的计算能力。

为了保障“云主站”在运行过程中的经济性、环境友好程度及可靠性,本文参考文献[17-18]刻画每一时段“云主站”的计算能耗,并将所有时段的总能耗作为目标,如式(17)所示。

$$g_2 = \sum_{r=1}^R \sum_{t=1}^T (A'u_r^\theta(t) + B) \quad (17)$$

其中, A' 、 B 、 θ 为能耗系数,根据文献[17-18]可知 $A' > 0$ 、 $B > 0$ 、 $\theta \in [2.5, 3]$ 。

3.1.2 约束条件

多目标优化问题的约束条件主要包括M/M/1排队模型约束、服务容忍时间约束与VM计算能力约束。M/M/1排队模型约束由式(16)所示响应延时为非负推导得到,具体如式(18)所示;服务容忍时间约束表示每台VM的平均响应延时须不大于应用最大容忍延时,如式(19)所示;VM计算能力约束表示每台VM配置的计算能力应当在容许的变化范围内,如式(20)所示。

$$\frac{u_r(t)}{L_r(t)} - \lambda_r(t) > 0 \quad r = 1, 2, \dots, R; t = 1, 2, \dots, T \quad (18)$$

$$0 \leq d_r(t) \leq D_r \quad r = 1, 2, \dots, R; t = 1, 2, \dots, T \quad (19)$$

$$u_{\min} \leq u_r(t) \leq u_{\max} \quad r = 1, 2, \dots, R; t = 1, 2, \dots, T \quad (20)$$

其中, D_r 为第 r 类业务应用的最大容忍延时; u_{\max} 和 u_{\min} 分别为VM计算能力的上、下限。

3.2 模型求解

3.2.1 多目标处理

为了统一解决响应延时最小和“云主站”能耗最小的min-min问题,本文先通过分段线性化方法对优化模型进行处理,然后采用归一化方法对2个目标进行处理。对于凸优化问题,当多目标权重构成的权重向量为非负时,归一化方法可得到帕累托前沿(PF)上的最优解^[19]。

为了统一处理2个度量标准不一致的目标,构造线性归一化函数对各目标进行归一化处理:

$$\varphi_j = \begin{cases} 1 & g_j \leq g_j^{\min} \\ (g_j - g_j^{\min}) / (g_j^{\max} - g_j^{\min}) & g_j^{\min} < g_j < g_j^{\max} \\ 0 & g_j \geq g_j^{\max} \end{cases} \quad (21)$$

其中, φ_j 为第 j 个目标函数归一化后的值, $j = 1, 2, \dots, N_{\text{obj}}$, N_{obj} 为目标函数的总数,本文中 $N_{\text{obj}} = 2$; g_j 为第 j 个目标函数值; g_j^{\max} 、 g_j^{\min} 分别为第 j 个目标函数的最大值、最小值。

基于归一化方法,可将原始模型转换为:

$$\begin{cases} \min \omega_1 \varphi_1 + \omega_2 \varphi_2 \\ \text{s.t. 式(15)---(20)} \end{cases} \quad (22)$$

其中, ω_1 、 ω_2 为两目标函数的权重,满足 $0 \leq \omega_1 \leq 1$, $\omega_2 = 1 - \omega_1$ 。通过遍历权重求得PF上足够多的最优解。

3.2.2 问题求解流程

步骤1:给定步长,遍历权重,采用MATLAB/CPLEX求解云资源的优化配置模型,得到足够多的PF最优解,构成PF最优解集合 X 。

步骤2:根据式(23)计算各目标函数的信息熵值 H_j ,并进一步根据式(24)获得各目标函数的权重。

$$H_j = - \sum_{i=1}^{N_p} \frac{h_{ij} \ln h_{ij}}{\ln N_p} \quad (23)$$

$$\omega_j = (1 - H_j) / \left(N_{\text{obj}} - \sum_{z=1}^{N_{\text{obj}}} H_z \right) \quad (24)$$

其中, N_p 为PF最优解总数; $h_{ij} = \varphi_{ij} / \sum_{i=1}^{N_p} \varphi_{ij}$, φ_{ij} 为第 j 个目标函数第 i 个PF最优解的值。

步骤3:采用TOPSIS通过相对接近度表征样本点与理想点之间的距离,以此作为选择集合 X 综合最优解的依据。

4 算例分析

为了验证本文所提方法的正确性,采用改进的IEEE 33节点多微电网系统作为测试系统,系统拓扑结构见附录中图A1。系统的相关参数见附录中表A1,综合能源应用功能设置见附录中表A2,计算负荷的相关参数见附录中表A3^[17,20]。

算例以各微电网的电气量数据为主要的运行数据,通过蒙特卡罗方法生成5000个样本场景。每个样本包含了各微电网与主网间联络线的功率、各微电网之间的交易功率、可再生能源计划及实际发电功率、电力用户的用电功率等24h的运行数据。借助K-均值聚类方法对5000个样本场景进行分类处理,并将24h的运行数据分为12个状态,假设每个状态对应1个事件,根据表A2中应用功能的设置确定每个事件对应的业务应用,其中某个样本场景的计算负荷见附录中图A2(a)。对经过分类处理的5000个样本场景的各时刻运行状态进行统计,获得初始状态概率矩阵、状态转移矩阵等,根据2.2节中状态关联的计算负荷模型,获取该测试系统的各类业务应用随着事件变化的计算负荷需求曲线。在计算负荷的建模阶段,式(12)和式(13)在电力物联网业务应用的固有计算量和固有到达频率的基础上,融入基于概率的电网运行状态时间分布特征。基于对不同电网运行状态计算负荷情况的考虑,生成以下3种计算负荷修正方法。

方法1:综合考虑各种运行状态下计算负荷的资源配置方法,修正系数 γ 均取值为1。

方法2:仅考虑最大概率状态下计算负荷的资源配置方法,每类业务应用在最大概率运行状态下的修正系数为1,在其余状态下的修正系数为0。

方法3:仅考虑各类业务应用计算负荷上边界的资源配置方法,修正公式如式(25)所示。

$$\begin{cases} \hat{L}_r(t_k) = \max \{ I_{L_r}(t_0), I_{L_r}(t_1), \dots, I_{L_r}(t_m) \} \\ \hat{\lambda}_r(t_k) = \max \{ I_{\lambda_r}(t_0), I_{\lambda_r}(t_1), \dots, I_{\lambda_r}(t_m) \} \end{cases} \quad (25)$$

基于上述3种计算负荷修正方法得到的计算负荷时间分布情况分别见附录中图A2(b)—(d)。其中,APP₁和APP₄的平均到达频率持续且恒定,这2类业务应用的计算负荷分布稳定,这是因为监测类和分析类业务应用需要不间断周期运行,系统规模在短时间内不会变化;APP₆和微电网与主网间的交易行为有关,微电网通过主网实现供需平衡,因此该类业务应用的计算负荷也基本上呈现恒定状态;APP₅与微电网间的交易行为有关,集中分布在00:00—08:00和18:00—24:00时段内,而APP₂、APP₃与风/光的实际和计划出力有关,通过向下控制命令调节实际出力以实现供需平衡,集中分布在

可再生能源出力较多的时段(08:00—18:00),方法1、2的计算负荷模型均能获得这3类业务应用的实际分布特点,而方法3考虑了各类业务应用的最大计算负荷,所得结果总体上比方法1、2的计算负荷大。对比方法1、2的结果可发现,考虑各种状态概率的业务应用计算负荷情况时,分布在08:00—18:00时段的计算负荷更大,综合考虑各种状态的可能性虽然在一定程度上增大了负荷值,但是能囊括低概率状态对计算负荷的影响,提高后续的应对能力。

针对上述3种计算负荷修正方法,对云计算资源配置情况进行优化,分别对应配置方法a—c。由于当前主站计算系统采用“一对一”(O2O)的资源配置模式^[9],即1类业务应用固定配置1台服务器,因此本文增加对照配置方法d以反映O2O配置模式:每类业务应用均配置1台服务器,保持服务器持续工作,按照最大计算能力配置计算资源。

采用上述4种资源配置方法进行资源优化配置后24h的响应延时和能耗结果见附录中图A3,总响应延时和总能耗结果见表2。可知,采用配置方法b时“云主站”的总能耗最小,其次为配置方法a、c,总能耗最大的为配置方法d;采用配置方法d时总响应延时最短,其次为配置方法c、a,最大的为配置方法b。这是因为配置方法d对每类业务应用均配置1台服务器,服务器不考虑虚拟技术,均以最大计算能力运行,因此业务应用的总处理延时会降低,但由于部分业务应用的计算量较小,且存在计算负荷时间分布特征,而配置方法d按照服务器的最大计算能力配置资源且服务器持续运行,总能耗会大幅增加。配置方法b虽然能够大幅降低“云主站”的总能耗,但资源仅能满足大概率状态下计算负荷的需求。与最大概率状态下的计算负荷相比,当概率较小状态下的计算负荷较大时,配置方法b配置的计算资源无法应对这些低概率状态出现时计算负荷陡增的情况,因此会出现“云主站”对部分业务应用响应延时较大甚至无法满足应用延时需求的情况,如在15:00—16:00时段内APP₅的延时较大。配置方法c按照各类业务应用的最大需求配置资源,能在一定程度上按照计算负荷需求分配,因此与配置方法b和a相比,其延时总体较好,但由于部分业务应用的计算负荷变化较大,无法按照计算负荷时间分布特征配置资源,必然也伴随着能耗增大的情况,所以

表2 4种配置方法的总响应延时和总能耗

Table 2 Total response time delay and total energy consumption of four collocation methods

配置方法	总响应延时/s	总能耗/(kW·h)	配置方法	总响应延时/s	总能耗/(kW·h)
a	93.849	18995.447	c	92.757	21620.884
b	95.945	18413.264	d	80.729	32593.445

其能耗最大严重制约了电力物联网绿色、经济、持续发展。结合表2结果可知,采用配置方法a能根据计算负荷的变化趋势合理配置资源,与配置方法b相比,其总能耗增加了3.16%,但能满足所有业务应用的延时要求,总延时减小了2.18%;与配置方法c相比,其总延时基本没变化,但总能耗减小了13.8%。

调整APP₁—APP₇的计算量参数,假设其计算量相同,平均响应延时和能耗随着计算量(CPU周期数)的变化曲线如图2所示。由图2可知,业务应用的平均响应延时、能耗均与计算负荷正相关。随着计算量增加,业务应用对“云主站”计算负荷的需求也增加,必然导致平均响应延时增加,为了满足业务应用的延时需求,配置的云资源计算速率加快,导致能耗增加。由图A2(a)和图2可知,APP₁、APP₄、APP₆这类持续执行的业务应用的平均到达频率较大,在24h内的平均延时及平均能耗较高,随着计算量增加,其平均响应延时和平均能耗增长更快;APP₂、APP₃集中分布在某一时段,在24h内的平均响应延时和平均能耗较小,随着计算量增加,其平均响应延时和平均能耗增长较缓。

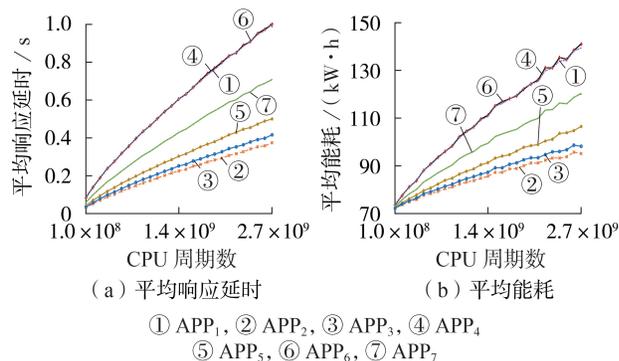


图2 平均响应延时和平均能耗随计算量的变化曲线

Fig.2 Curves of average response time delay and energy consumption vs. computation load

5 结论

本文基于云计算架构的自动化系统新模式,考虑业务应用与电网运行状态之间的强关联,提出了电力物联网计算负荷的建模方法,并将其应用于计算资源动态优化配置,结合测试系统的历史运行数据及业务应用功能部署对多个场景进行仿真分析,得到以下结论:

(1)基于电网运行特性,建立基于马尔科夫链的电网运行状态模型,并根据电力物联网的业务处理流程梳理典型电力物联网业务应用及业务逻辑,总结状态-事件-应用关联关系,实现状态关联的计算负荷表征;

(2)在电力物联网业务应用固有特征的基础上,融入基于概率的电网运行状态特征,将其修正至每

一时刻业务应用的计算负荷总量估计中,算例结果表明综合考虑各种状态的计算负荷能够准确掌握计算负荷的变化特性;

(3)通过对比不同计算负荷修正方法以及当前主流O2O资源分配模式,验证了综合考虑运行状态概率的计算负荷模型能够为资源配置提供依据,在主站层面能够满足响应延时需求及减少主站能耗。

附录见本刊网络版(<http://www.epae.cn>)。

参考文献:

- [1] 董朝阳,赵俊华,文福拴,等.从智能电网到能源互联网:基本概念与研究框架[J].电力系统自动化,2014,38(15):1-11. DONG Zhaoyang,ZHAO Junhua,WEN Fushuan,et al. From smart grid to energy internet:basic concept and research framework[J]. Automation of Electric Power Systems,2014,38(15):1-11.
- [2] 王东芳,刘水源,张勇军,等.“互联网+”形势下电力信息物理融合发展研究综述与展望[J].电力自动化设备,2020,40(6):90-99. WANG Dongfang,LIU Shuiyuan,ZHANG Yongjun,et al. Review and prospect of cyber-physical development of power system under background of “Internet+” technology[J]. Electric Power Automation Equipment,2020,40(6):90-99.
- [3] 王相伟,史玉良,张建林,等.基于Hadoop的用电信息大数据计算服务及应用[J].电网技术,2015,39(11):3128-3133. WANG Xiangwei,SHI Yuliang,ZHANG Jianlin,et al. Computation services and applications of electricity big data based on Hadoop[J]. Power System Technology,2015,39(11):3128-3133.
- [4] 岳青鹏,胡志坚,陈纬楠.物联网环境下考虑多个利益主体的区域配电系统能量管理[J].电力自动化设备,2020,40(8):111-118. YUE Jingpeng,HU Zhijian,CHEN Weinan. Energy management of regional distribution system considering multiple stakeholders under IoT environment[J]. Electric Power Automation Equipment,2020,40(8):111-118.
- [5] 杨鹏,申洪涛,陶鹏,等.云平台下时间序列数据并行化排列熵特征提取方法[J].电力自动化设备,2019,39(4):217-223. YANG Peng,SHEN Hongtao,TAO Peng,et al. Parallel permutation entropy feature extraction method for time series data based on cloud platform[J]. Electric Power Automation Equipment,2019,39(4):217-223.
- [6] 应俊,蔡月明,刘明祥,等.适用于配电物联网的低压智能终端自适应接入方法[J].电力系统自动化,2020,44(2):22-27. YING Jun,CAI Yueming,LIU Mingxiang,et al. Adaptive access method of low voltage intelligent terminal for distribution Internet of Things[J]. Automation of Electric Power Systems,2020,44(2):22-27.
- [7] BENNANI M N,MENASCE D A. Resource allocation for autonomic data centers using analytic performance models[C]// International Conference on Autonomic Computing. Seattle, WA, USA:IEEE,2005:229-240.
- [8] CALHEIROS R N,MASOUMI E,RANJAN R,et al. Workload prediction using ARIMA model and its impact on cloud applications’ QoS[J]. IEEE Transactions on Cloud Computing,2015,3(4):449-458.
- [9] CHEN Z Y,HU J,MIN G Y,et al. Towards accurate prediction for high-dimensional and highly-variable cloud workloads with deep learning[J]. IEEE Transactions on Parallel and Distributed Systems,2020,31(4):923-934.

- [10] MASOUIROS D,XYDIS S,SOUDRIS D. Rusty:runtime system predictability leveraging LSTM neural networks[J]. IEEE Computer Architecture Letters,2019,18(2):103-106.
- [11] GUNJAL P R,GUNJAL B R,SHINDE H A, et al. Moving object tracking using Kalman filter[C]//2018 International Conference on Advances in Communication and Computing Technology (ICACCT). Sangamner,India:IEEE,2018:544-547.
- [12] XIONG K,PERROS H G. Service performance and analysis in cloud computing[C]//Congress on Services-I. Los Angeles, CA,USA:IEEE,2009:693-700.
- [13] STREHL K,MORAGA C,TEMME K, et al. Fuzzy decision diagrams for the representation, analysis and optimization of rule bases[C]//Proceedings of the 30th IEEE International Symposium on Multiple-Valued Logic (ISMVL 2000). Portland, OR,USA:IEEE,2000:127-132.
- [14] PRADO R P,GARCÍA-GALÁN S,MUÑOZ EXPÓSITO J E, et al. Knowledge acquisition in fuzzy-rule-based systems with particle-swarm optimization[J]. IEEE Transactions on Fuzzy Systems,2010,18(6):1083-1097.
- [15] FAN Q,ANSARI N. Application aware workload allocation for edge computing-based IoT[J]. IEEE Internet of Things Journal,2018,5(3):2146-2153.
- [16] WANG Y,YANG R,WO T, et al. Improving utilization through dynamic VM resource allocation in hybrid cloud environment [C]//2014 20th IEEE International Conference on Parallel and Distributed Systems. Hsinchu, Taiwan, China:IEEE,2014: 241-248.
- [17] DENG R L,LU R X,LAI C Z, et al. Optimal workload allocation in fog-cloud computing towards balanced delay and power consumption[J]. IEEE Internet of Things Journal,2016, 3(6):1171-1181.
- [18] GUO M,LI L,GUAN Q S. Energy-efficient and delay guaranteed workload allocation in IoT-edge-cloud computing systems[J]. IEEE Access,2019,7:78685-78697.
- [19] 陈聪,沈欣炜,夏天,等. 计及效率的综合能源系统多目标优化调度方法[J]. 电力系统自动化,2019,43(12):60-67,121. CHEN Cong, SHEN Xinwei, XIA Tian, et al. Multi-objective optimal dispatch method for integrated energy system considering exergy efficiency[J]. Automation of Electric Power Systems,2019,43(12):60-67,121.
- [20] REN J K,YU G D,HE Y H, et al. Collaborative cloud and edge computing for latency minimization[J]. IEEE Transactions on Vehicular Technology,2019,68(5):5031-5044.

作者简介:



孙宇嫣

孙宇嫣(1995—),女,广东中山人,硕士研究生,主要研究方向为电力物联网、能源互联网(**E-mail**:894597048@qq.com);

蔡泽祥(1960—),男,江苏南京人,教授,博士研究生导师,主要研究方向为电力系统保护、控制与自动化(**E-mail**:epzxcai@scut.edu.cn);

马国龙(1993—),男,河南洛阳人,硕士研究生,主要研究方向为微电网运行与控制、能源互联网(**E-mail**:1185849557@qq.com)。

(编辑 陆丹)

Workload model and optimal resource allocation of cloud master station in Power Internet of Things

SUN Yuyan, CAI Zexiang, MA Guolong, GUO Caishan

(School of Electric Power, South China University of Technology, Guangzhou 510640, China)

Abstract: With the continuous development of PloT (Power Internet of Things), business applications such as operation analysis, energy trading, user energy efficiency management, power management, and so on, show exponential growth, which brings huge computing and processing pressure to the traditional master station system. In order to solve the above problems, the workload model of PloT is established for the new generation of cloud master station architecture, and based on this, a cloud resource allocation method is proposed. Starting from tracing the root of the dynamic change of workload in the cloud master station, the change rule of power grid operation state is depicted by using Markov chain. According to the specification of business processing flow, the state-related workload model is established, including the workload description model and the state-event-application correlation model, which is established as an experience knowledge base for analyzing the workload characteristics of various businesses in different periods. Based on the above workload model, a dual-objective optimization model for cloud resource allocation is established with the goals of shortest average response time delay and minimum energy consumption of cloud master station. The comprehensive optimal solution is obtained from the Pareto frontier solution set by using TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution). Taking the modified IEEE 33-bus system as an example, simulative results verify that the proposed method can effectively analyze the variation characteristics of workload in different applications, and then reasonably allocate cloud resource.

Key words: PloT; cloud master station; workload model; cloud resource allocation; Markov chain; experience knowledge base; TOPSIS

附录

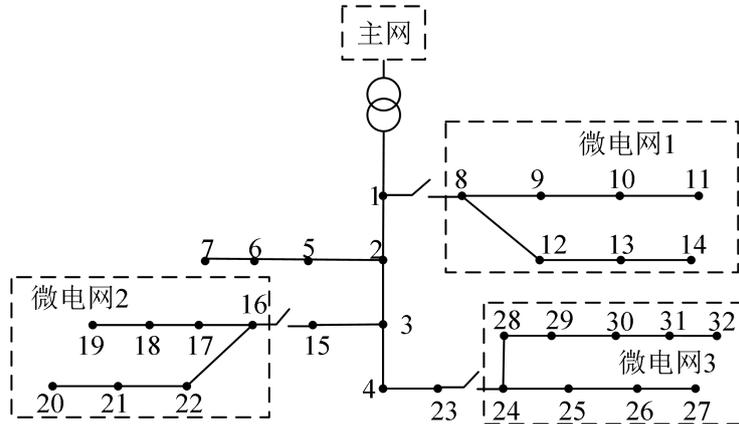


图 A1 改进的 IEEE 33 节点系统

Fig.A1 Modified IEEE 33-bus system

表 A1 系统元件参数

Table A1 Parameters of components in system

微电网	光伏额定功率/kW	风机额定功率/kW	储能额定功率/kW	储能额定容量/(kW·h)
1	200	500	150	300
2	300	600	200	500
3	500	300	200	500

表 A2 应用的相关参数

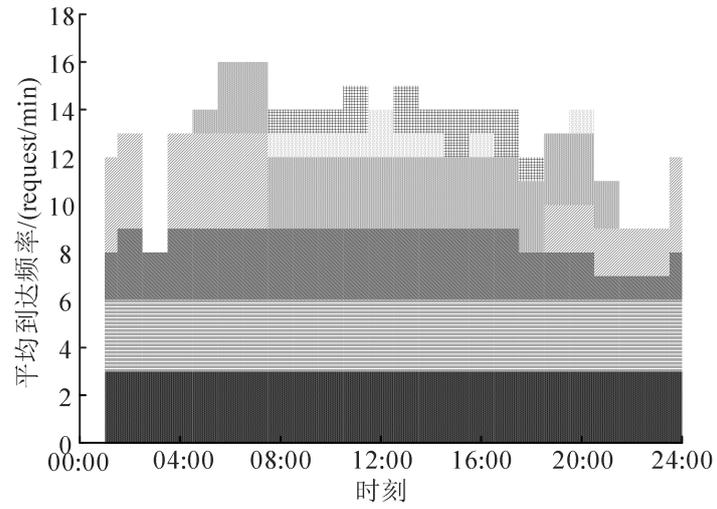
Table A2 Related parameters of application

应用类别	编号	功能描述
状态监测	APP ₁	数据采集、状态估计
发电控制	APP ₂	光伏发电控制
	APP ₃	风机发电控制
运行分析	APP ₄	拓扑分析、潮流计算、安全评估
电力交易	APP ₅	微网间交易
	APP ₆	微网与主网交易
能量管理	APP ₇	能量管理、储能策略

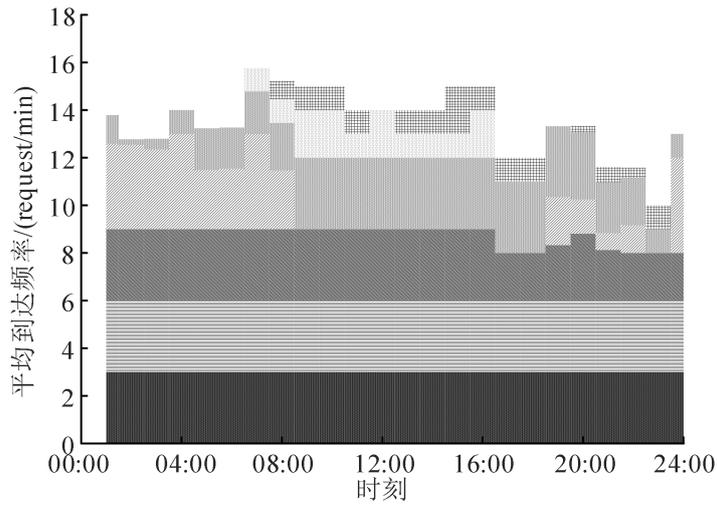
表 A3 计算负荷的相关参数

Table A3 Related parameters of workload

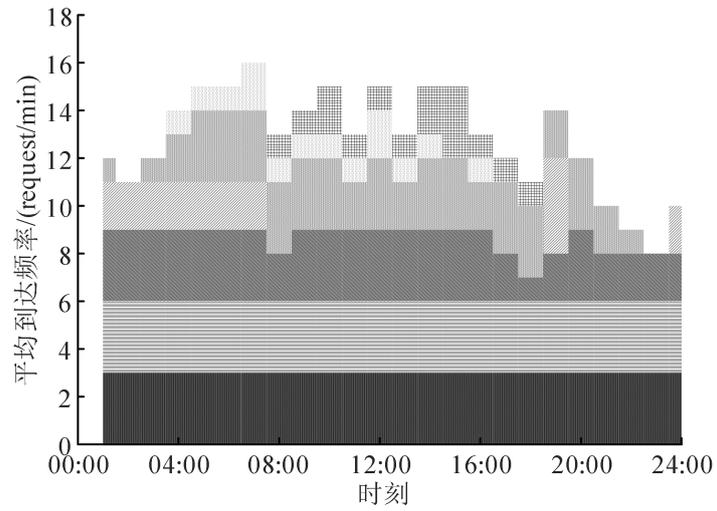
参数	取值
CPU 周期数	服从 $[10^9, 3 \times 10^9]$ 内的均匀分布
微电网内每个应用请求到达时间	1 min
能耗参数	$A' = 3.206 \times 10^{-27}, B = 68, \theta = 3$
“云主站”计算能力	$u_{\min} = 1.0 \text{ GHz}, u_{\max} = 3.4 \text{ GHz}$
最大容忍响应时间	1.0 s



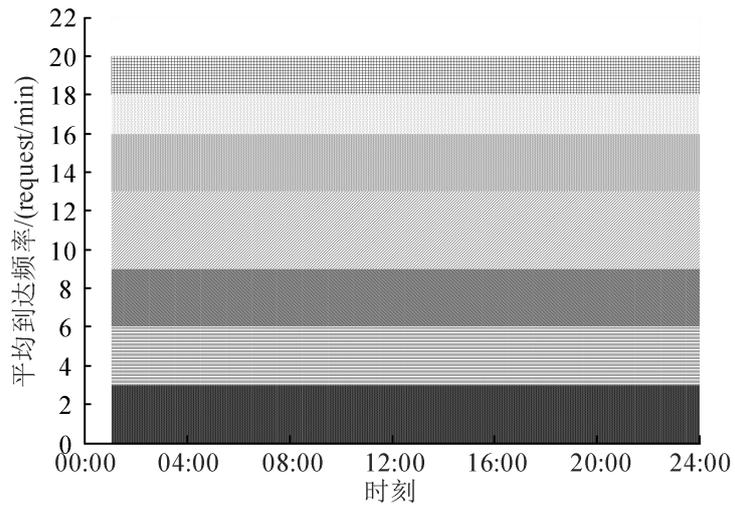
(a) 测试样本



(b) 方法 1



(c) 方法 2

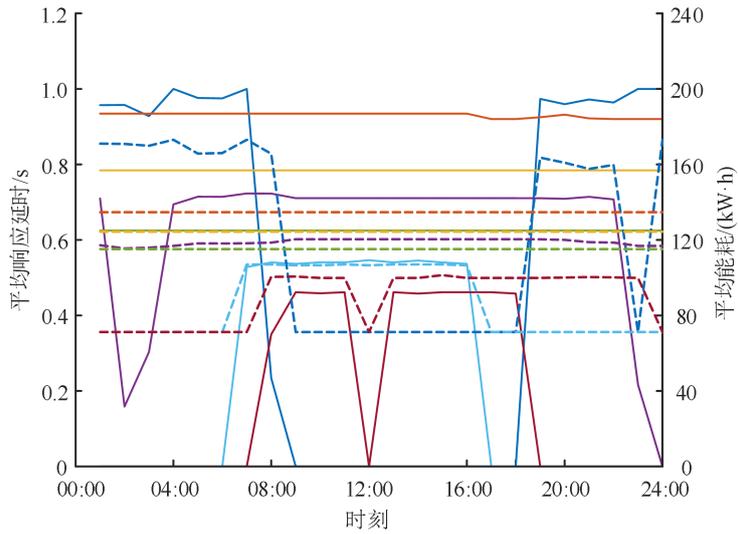


(d) 方法 3

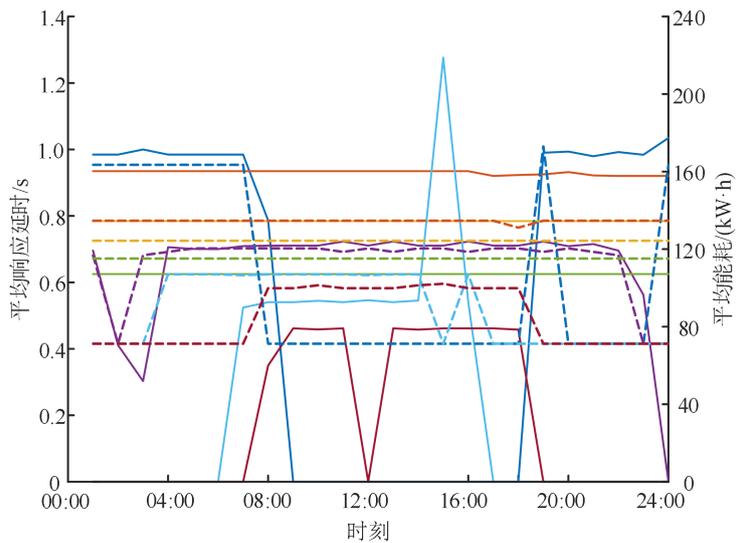
■ APP₁ □ APP₂ ▨ APP₃ ▩ APP₄ ▪ APP₅ ▫ APP₆ ▬ APP₇

图 A2 计算负荷的时间分布

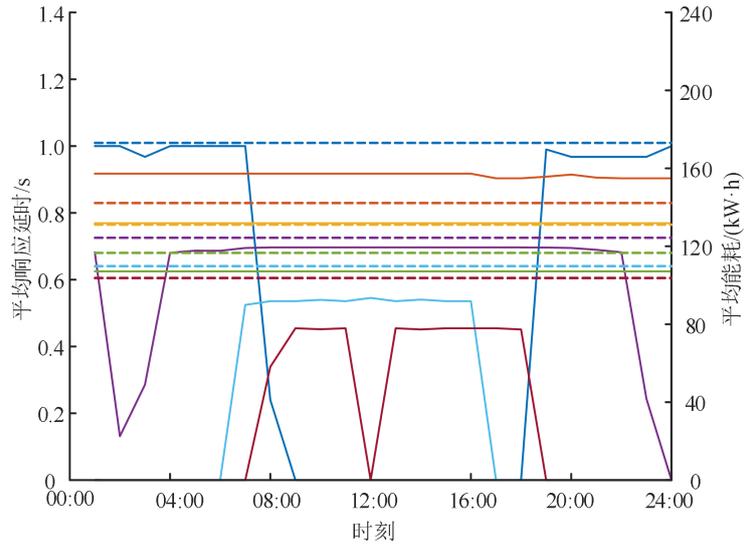
Fig.A2 Time distribution of workloads



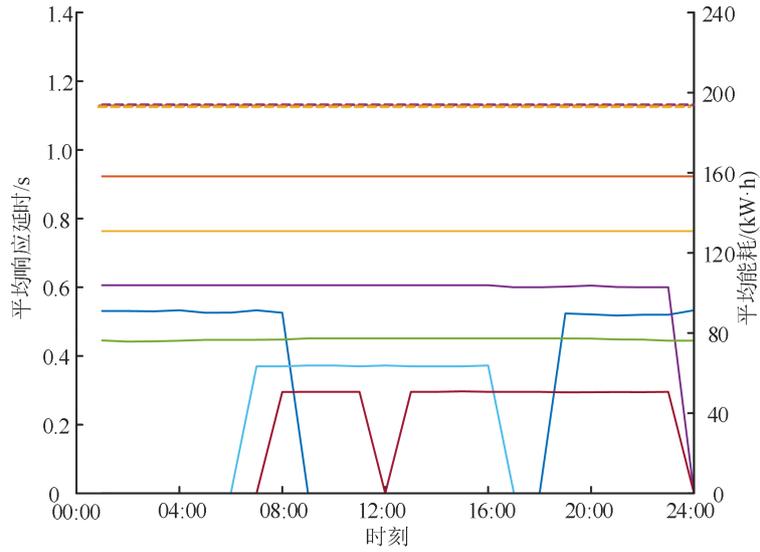
(a) 分配方法 a



(b) 分配方法 b



(c) 分配方法 c



(d) 分配方法 d

延时 — APP₁ — APP₂ — APP₃ — APP₄ — APP₅ — APP₆ — APP₇
 能耗 - - APP₁ - - APP₂ - - APP₃ - - APP₄ - - APP₅ - - APP₆ - - APP₇

图 A3 4 种分配方法的平均响应延时和平均能耗

Fig.A3 Average response time delay and average energy consumption of four allocation methods