

# 基于性能量化矩阵的计算网格作业调度算法研究<sup>\*</sup>

黄义番<sup>1</sup> 郭庆平<sup>1</sup> 刘 鹏<sup>2</sup> 晏志超<sup>1,2</sup>

(武汉理工大学 武汉 430063)<sup>1</sup>

(中国人民解放军理工大学全军军事网格研究中心 南京 210007)<sup>2</sup>

**摘要** 提升计算网格系统运行效率的关键在于作业调度算法,如何综合各种因素使得调度策略更为全面是一个有挑战性的问题。通过建立网格资源性能量化矩阵,构建了一个作业调度模型,并基于此模型给出了一个具体的作业调度算法。通过性能分析和实验仿真,该算法在运行时间、占用资源等方面都有较大的改善,能较好地适应网格系统的动态性和可扩展性特点,提高了作业调度效率。

**关键词** 计算网格,网格资源,性能量化矩阵,信任服务质量

## Research on Job Scheduling Algorithm of Computing Grid Based on Performance Measured-Matrix

HUANG Yi-fan<sup>1</sup> GUO Qing-ping<sup>1</sup> LIU Peng<sup>2</sup> YAN Zhi-chao<sup>1,2</sup>

(School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430063, China)<sup>1</sup>

(PLA Research Center for Military Grid Technology, PLA University of Science and Technology, Nanjing 210007, China)<sup>2</sup>

**Abstract** The key technology to improve the grid system operating efficiency is the job scheduling algorithm, how to integrate various factors to make the scheduling algorithm more comprehensive is a challenging problem. By establishing grid resources performance measured-matrix, the paper constructed a job scheduling model, and gave a specific job scheduling algorithm based on this model. Through the performance analysis and the experimental simulation, this algorithm improved the executing time, took few resources, and better adapted to the dynamic and scalability characteristics, which enhanced the job scheduling efficiency.

**Keywords** Computing grid, Grid resources, Performance measured-matrix, Trust quality of service

## 1 引言

随着 Internet 技术的不断发展,将地理上广泛分布的大量计算资源组合起来进行大规模问题的求解变得日益普遍,由此产生了网格计算<sup>[1]</sup>。网格计算环境具有动态性、异构性、可扩展性及多级管理域等特性<sup>[2]</sup>。

正是由于计算网格结构的复杂性,使得网格计算环境中的作业调度问题一直是一个 NP 难解问题。大量的网格计算项目试图提供一个合适的作业调度方法,可是都存在一定程度的缺陷<sup>[3]</sup>。因此合理地作业进行调度,以科学地利用网格的各种资源便成为一项意义重大的课题。

本文将基于性能量化矩阵,提出一种新的作业调度模型来达到简化作业调度的目的。

## 2 相关工作

### 2.1 计算网格资源结构

计算网格环境中计算资源的组成结构十分复杂,它可以由若干 PC 机、服务器、集群计算机、SMP 高性能计算机、MPP 高性能计算机等构成<sup>[4]</sup>。如图 1 所示,网格环境中网格资源的组成包括 PC 机、集群、局域网、SMP 高性能计算机等,一个集群或者局域网可能同时也包括数个 PC 机、集群和局域网等资源。

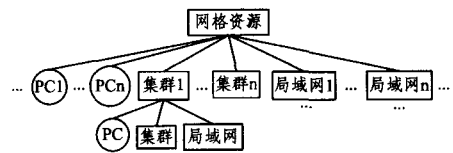


图 1 网格环境中计算资源的组成结构图

### 2.2 作业调度系统模型

根据网格计算环境的特性和用户作业提交的实际情况,作业的流程一般为用户提交作业,网格系统接收到作业,安排进入等待队列,等待作业调度器的调度,而作业调度器需要资源信息中心给它返回网格资源的信息,进而给作业分派资源来执行。作业调度系统模型结构如图 2 所示,图中的网格资源模块与图 1 中的网格资源是一个概念,代表相同的信息,即网格计算资源。

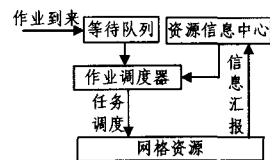


图 2 作业调度系统模型

<sup>\*</sup>基金项目:国家自然科学基金资助项目(批准号:60403043)。黄义番 硕士研究生,主要研究方向为分布式并行计算、网格计算;郭庆平 教授,博士生导师,主要研究方向为分布式并行处理、信息安全、网格计算;刘 鹏 副教授,博士,主要研究方向为网格计算。

### 3 作业调度性能量化矩阵的构建

基于网格资源的特性,我们从3个方面来构建性能量化矩阵,这3个方面包括网格资源的处理能力指标、信任 QoS 和作业调度资源时所需要考虑的效益。

#### 3.1 网格资源处理能力量化矩阵

网格中的各种资源存在较大区别,不同的资源不仅处理的任务不同,而且不同资源处理相同任务的能力也不同。而这些性能的差异,主要体现在资源节点配置的不同上,这就需要采取量化的方式描述资源的信息。

为简化资源处理能力矩阵的建立,我们采用静态的方式,主要考虑资源节点的几种主要配置参数:CPU 速度、资源构成、内存大小等。设  $x, y, z \dots$  分别表示资源节点的 CPU 速度,资源构成参数值和内存大小值,资源能力函数为  $f(x, y, z \dots)$ 。

则由资源能力函数可知资源  $k$  处理作业  $i$  的能力值为:

$$a_{ik} = \begin{cases} 0 & \sum x_i < 0 \\ f(x_{ik}, y_{ik}, z_{ik} \dots) & \sum x_i \geq 0 \end{cases}$$

其中  $x_{ik}, y_{ik}, z_{ik} \dots$  表示资源  $k$  的参数变量,  $x_i$  表示调度作业对资源某一方面的需求,  $\sum x_i$  表示作业对资源的总体需求。当  $\sum x_i < 0$  时,则表示资源不满足作业的最低执行权限,这样赋值为 0,表示资源不能执行该作业。

因此,  $m$  个作业组成的队列分配给  $n$  个可用资源的处理能力量化矩阵为:

$$A_{ability} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \quad (1)$$

#### 3.2 信任 QoS 量化矩阵

信任 QoS 是一个非常复杂的主观概念,目前也没有一致的定义,无法同资源配置一样可客观描述<sup>[5,6]</sup>。本文给出定义如下:

**定义 1** 信任 QoS 由信任值表征的客观实体的身份和行为的可信度评估,信任值取决于实体可靠性、诚信和性能。计算网格信任模型主要由资源信任属性、作业信任属性及其相互信任关系构成。

资源信任属性、作业信任属性及其相互间的信任关系本文统一描述,三者可概括为如下两个方面<sup>[7]</sup>:

1) 安全性:衡量网格资源中作业和数据的真实性、保密性和完整性的保障程度;

2) 可靠性:长时间执行的作业可能因为某个资源离开网格系统而导致运行中止,造成作业无法继续执行。

基于以上分析,我们采用信任函数  $h(t, r)$  来求解资源的信任 QoS 值,其中  $t, r$  是两个自变量,  $t$  表示资源的安全性,  $r$  表示资源的可靠性,信任函数  $h(t, r) = t \times r$ 。

则由信任函数可知资源  $k$  处理作业  $i$  的信任 QoS 值为:

$$b_{ik} = \begin{cases} 0 & \sum y_i < 0 \\ h(t_{ik}, r_{ik}) & \sum y_i \geq 0 \end{cases}$$

其中  $t_{ik}, r_{ik}$  分别表示资源  $k$  处理作业  $i$  的安全性和可靠性属性值,  $y_i$  表示调度作业对资源某一方面信任度的需求,  $\sum y_i$  表示作业对资源的总体信任度需求。当  $\sum y_i < 0$  时,则表示资源不满足作业的最低信任值需求,这样赋值为 0,表示资源不能执行该作业。

因此,  $m$  个作业组成的队列分配给  $n$  个可用资源的信任

QoS 量化矩阵为:

$$B_{trust} = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{pmatrix} \quad (2)$$

#### 3.3 效益量化矩阵

**定义 2** 效益是指作业在资源上的预测执行时间和作业从提交到执行完毕所付出的代价,包括时间和代价两个属性<sup>[8]</sup>。

基于以上定义,我们采用效益函数  $g(d, e)$  来求解资源处理作业的效益值,其中  $d, e$  是两个自变量,  $d$  表示效益的时间属性,  $e$  表示效益的代价属性,效益函数为:

$$g(d, e) = \frac{1}{d \times e}$$

则由效益函数可知资源  $k$  处理作业  $i$  的效益值为:

$$c_{ik} = g(d_{ik}, e_{ik}) = \frac{1}{d_{ik} \times e_{ik}}$$

其中  $d_{ik}, e_{ik}$  分别表示资源  $k$  处理作业  $i$  的时间属性值和代价属性值,因此,  $m$  个作业组成的队列分配给  $n$  个可用资源的效益量化矩阵为:

$$C_{benefit} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{pmatrix} \quad (3)$$

#### 3.4 作业调度系统性能量化矩阵

对于以上得出的 3 个矩阵,进行简单整合就可以求出作业调度系统所需的性能量化矩阵,用  $R_{m,n}$  表示性能量化矩阵,采用加法整合,即  $R_{m,n} = A_{ability} + B_{trust} + C_{benefit}$ 。

$$\text{令: } R_{m,n} = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \dots & \dots & \dots & \dots \\ d_{m1} & d_{m2} & \dots & d_{mn} \end{pmatrix} \quad (4)$$

其中性能量化矩阵的每一个分量  $d_{ik} = a_{ik} + b_{ik} + c_{ik}$ 。

通过矩阵整合,得出了网格作业调度系统的性能量化矩阵,下面进一步对量化矩阵进行分析,从而达到把复杂问题简单化的目的。  $R_{m,n}$  矩阵中的每一个分量都可以看作某一时刻资源对用户作业请求的满足值,分量  $a_{ik} + b_{ik} + c_{ik}$  反应出了资源  $k$  对作业  $i$  的满足程度,分量值越大说明处理作业的性能也越好。另外,由于网格资源的动态性特性和数据预测值的精确性不高,决定了不必得出  $R_{m,n}$  的精确值。因此,不妨使用  $e_{ik} = \sqrt[3]{a_{ik} \times b_{ik} \times c_{ik}}$  来求  $R_{m,n}$ ,即采用几何平均值求解量化值,从而得出最终的性能量化矩阵为:

$$R_{m,n} = \begin{pmatrix} e_{11} & e_{12} & \dots & e_{1n} \\ e_{21} & e_{22} & \dots & e_{2n} \\ \dots & \dots & \dots & \dots \\ e_{m1} & e_{m2} & \dots & e_{mn} \end{pmatrix} \quad (5)$$

### 4 基于性能量化矩阵的网格作业调度算法

#### 4.1 计算网格作业调度模型

由性能量化矩阵  $R_{m,n}$  的推导过程,计算出每个分量的具体值,并以分量的具体值作为作业调度模型中网格资源的权系数。下面给出基于性能量化矩阵建立的作业调度模型的层次结构,如图 3 所示。

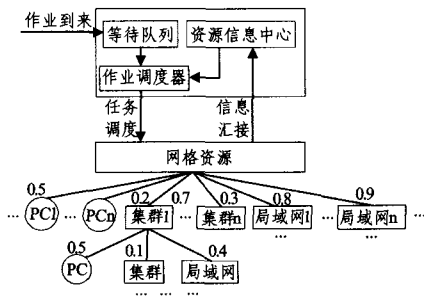


图3 基于量化矩阵的作业调度模型

图3中每个资源节点上面标出的数字即为性能量化矩阵每个分量的计算值 $e_{ik}$ ,性能量化矩阵的第 $i$ 行, $n$ 个分量组成的向量表示第 $i$ 个作业分配在 $n$ 个可供调度的资源上的性能值。某一时刻作业调度模型中的作业调度器就可以根据这一权重系数值来进行资源调度,进而分配资源给作业。

#### 4.2 基于量化调度模型的作业调度算法

考虑到量化调度模型的层次结构特点,作业调度算法依然采用循环控制和条件判断的步骤进行,只是算法中融入了性能量化矩阵,使算法更加简单且易于实现,算法步骤如下:

Measured-Matrix 算法 //算法名字描述

Step1

输入:资源集合: $R=\{r_1, r_2, r_3, \dots, r_n\}$

作业集合: $T=\{t_1, t_2, t_3, \dots, t_m\}$

Step2

对 $T$ 中的作业根据其资源的请求信息排入等待队列;

$i=1$ ; //当前作业的下标初始化为1

Step3

while( $T$ 中有作业未完成)do

    计算出资源调度的性能量化矩阵 $R_{m,n}$ ;

    while( $R$ 中有资源未利用)do

        if(第 $i$ 个作业已完成)then

$i=i+1$ ;

        if( $i>m$ )then

$i=1$ ;

        End if

    End i

    将作业 $i$ 提交给 $R$ 中第一个未利用的资源(记为 $k$ );

$i=i+1$ ;

    If( $i>m$ )then

$i=i+1$ ;

    End if

    End while

接收各个完成的作业,记录其完成的资源号和完成时间;

对完成作业的资源 $k$ (设其完成的作业为 $i$ )再次加入网络空闲资源;

End while

Measured-Matrix 算法分3步执行:

第一步:首先输入网络系统中当前条件下的资源集 $R$ 和作业队列 $T$ ;

第二步:初始化作业队列,并把当前即将被调度的作业的下标初始化为1;

第三步:是算法最重要的一步,包括两个循环,一个是对作业的循环,还有一个是对资源的循环,通过作业循环不断更新量化矩阵的值,进而达到作业-资源的最佳匹配。首先计算出作业1被部署到 $n$ 个资源的性能量化值,然后调度最佳

资源,分配当前作业;接着计算出第2个作业调用 $n$ 个资源的性能量化值,然后调度最佳资源,分配当前作业。依此循环下去,直到作业集分配完成。另外,该步骤在给每个作业寻找计算资源的时候,同时也回收资源,这一特点能很好地适应网络资源的动态特性和可扩展性。

## 5 算法性能分析

为了充分说明 Measured-Matrix 算法的正确性和体现其优越性,我们从两个方面对该作业调度算法进行了分析:复杂度和仿真分析。

### 5.1 复杂度分析

#### 5.1.1 空间复杂度分析

该算法在执行之前,必须为每一个资源节点分配空间,用来存储各自的性能指标,在前面调度模型建立的过程中已假设有 $n$ 个资源,如果假设每个资源占用的空间为 $s_i$ ,则静态分配给资源节点的空间应为 $S=\sum_{i=1}^n s_i$ 。另外,该算法在执行的过程中会根据作业的提交,计算出作业调度过程中所需要的性能量化矩阵,因此应动态地提供适当的空间,以保存计算结果存储。设量化矩阵一个分量占用的空间为 $d$ ,则 $m \times n$ 阶矩阵占用的空间为 $\sum_{i=1}^m m \times d$ ,因此 Measured-Matrix 算法执行过程中所需要的空间应为 $\sum_{i=1}^m (s_i + m \times d)$ 。

#### 5.1.2 时间复杂度分析

基于量化矩阵的 Measured-Matrix 算法,首先要求出调度作业的性能量化矩阵。前面已假设有 $n$ 个资源,则每一个作业调度最佳资源时需要计算 $n$ 次。然后判断资源是否空闲且满足作业调度条件,这一过程也需要 $n$ 次。另外,作业等待队列的长度为 $m$ ,因此该算法的时间复杂度应为 $O(m \times n)$ 。

### 5.2 仿真分析

为了更好地检验 Measured-Matrix 算法的应用效果,我们使用 SimGrid 模拟器<sup>[9]</sup>对所提出的算法进行仿真研究。因为 SimGrid 模拟器提供了一系列 API,可以方便地随机生成不同的主机处理能力、网络带宽、数据传输量和计算量等参数。

考虑到网络计算环境中的网络资源的处理能力、信任函数值和信任 QoS 等都不同,因此在进行作业调度和处理的过程中,必须综合考虑这些因素来求出性能量化矩阵,进而调度作业。我们假定作业调度都是通过中央调度器,初始化网络资源个数 $n$ 为5,各个资源的性能指标初始化值见表1所示。

表1 初始化网络资源的性能指标

因素	处理能力	信任 QoS	效益值
集群	4.0	1.0	0.3
局域网	1.5	1.3	0.2
PC1	0.8	2.0	1.0
PC2	0.5	0.8	1.1
PC3	0.5	3.0	0.6

根据网络资源的初始化性能指标值,我们由性能量化矩阵(5)的数学推导公式即可求出性能量化矩阵。图4是 Measured-Matrix 算法在网络资源数量 $n$ 为固定值5,且性能指标为表1所提供的情况下,不同数量的作业提交到网络系统时得到的作业平均完成时间和作业提交量的关系。

(下转第132页)

用率之间找到合适的关系。因此,Web 服务需要 QoS 的保证,服务质量评估也就成为电子商务中的重要部分。本文既考虑用户得出评价的上下文,又将相似期望连同上下文一起考虑,使得评价更具有实际意义。针对服务质量评估特性,建模中采用模糊数学方法,提出了一个计算服务 QoS 的扩展模型。在评价当中,既对用户的偏好和个性化观点加以考虑,又根据用户的地位和声誉确定用户评价的重要程度,因此所得计算结果更加精确而有实际意义。

### 参考文献

[1] Schillo M, Funk P, Rovatsos M. Using trust for detecting deceitful agents in artificial societies. *Applied Artificial Intelligence, Special Issue on Trust, Deception and Fraud in Agent Societies*, 2000; 825-848

[2] Yu B, Singh M. An evidential model of distributed reputation management // *Proceedings of First International Conference on Autonomous Agents and Multi Agent Systems*, 2002; 294-301

[3] Deora V, Shao J, Gray W A, et al. A quality of service management framework based on user expectations // *Proceedings of the*

*First International Conference on Service Oriented Computing (ICSOC)*. Italy, December 2003; 104-114

[4] Deora V, Shao J, Gray W A, et al. Expectation-based quality of service assessment. *International Journal on Digital Libraries*, 2006

[5] Deora V, Shao J, Gray W A, et al. Modelling quality of service in service oriented computing // *Proceedings of the Second IEEE International Symposium on Service-oriented System Engineering (SOSE)*, 2006

[6] 陈彦萍, 李增智, 等. Web 服务中基于服务质量的服务选择算法. *西安交通大学学报*, 2006

[7] 谢季坚, 刘承平. 模糊数学方法及其应用. 华中科技大学出版社, 2006

[8] Yang L, Dai Y, Gao Y. A service oriented web services composition platform. *Journal of Wuhan University*, 2006, 11(1): 160-164

[9] Liao Y, Tang L, Li M S. A method of QoS-aware service component composition [J]. *Chinese Journal of Computers*, 2005, 28(4): 627-634

(上接第 78 页)

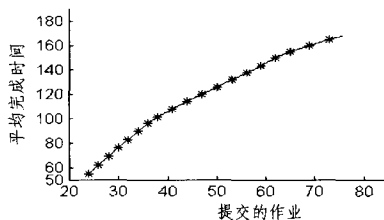


图 4 相同网格资源情况下的作业平均完成时间

从以上算法性能分析中可以看出,基于量化作业调度模型的算法具有较小的空间复杂度和时间复杂度,而且具有较好的应用效果。仿真结果(图 4)反映出了随着作业数量的增加,作业平均完成时间也相应增加,两者几乎成正比例。因此,它是一个较为合理的高效作业调度算法。而且由于 Measured-Matrix 算法引入了性能量化矩阵,因此能更好地适应网格计算的特性。

**结束语** 一个好的调度算法可以提高系统资源的利用率和吞吐量,缩短作业的响应时间,从而提供不间断的高效的计算服务。本文正是通过资源性能预测,计算出作业调度资源的性能量化矩阵,量化矩阵的建立有效地降低了调度的复杂性,提高了作业调度的批量能力,该算法具有运行效率高、占用资源少的特点。同时,由于调度算法是基于量化网格作业调度模型的,这一特点可很好地适应网格系统中资源的动态特性和可扩展特性。然而,由于网格计算环境中资源和作业信息的不确定性和资源管理上的独立性,很难估算出性能量化矩阵的计算前提。这一难题将是今后继续研究的课题,本文的量化层次调度模型和 Measured-Matrix 算法可为今后这

一课题提供一定的参考。

### 参考文献

[1] Intel Corporation. Intel Itanium 2 Processor Reference Manual for Software Development and Optimization [EB/OL]. <http://developer.intel.com/design/itanium2/manuals/index.htm>. 2002, 6

[2] 王汝传, 韩光法. 网格计算环境下作业调度的策略研究[J]. *重庆邮电学院学报*, 2005, 17(2): 198-203

[3] 徐志伟, 冯百明, 李伟. 网格计算技术[M]. 北京: 电子工业出版社, 2004; 229-235

[4] 张树东, 曹元大. 机群环境中基于信度模型的作业调度算法[J]. *北京航空航天大学学报*, 2004, 30(11): 1097-1081

[5] Scherwani J, Ali N, Lotia N, et al. Libra: A Computational Economy-based Job Scheduling System for Clusters [J]. *Software Practice and Experience*, 2004, 6(11): 581-598

[6] Grandison T, Sloman M. A survey of trust in internet applications[J]. *IEEE Communications Surveys & Tutorials*, 2000, 3(4): 2-16

[7] 张伟哲, 方滨心, 胡铭曾, 等. 基于信任 QoS 增强的网格服务调度算法[J]. *计算机学报*, 2006, 29(7): 1158-1165

[8] Fujimoto N, Hagihara K. A Comparison Among Grid Scheduling Algorithms for Independent Coarse-Grained Tasks // *SAINT 2004 Workshop on High Performance Grid Computing and Networking*. IEEE Press, January 2004; 674-680

[9] Casanova H. Simgrid: A Toolkit for the Simulation of Application Scheduling[A] // *Proceedings of the 1st International Symposium on Cluster Computing and the Grid[C]*. 2001; 430