

# 基于服务质量的网格 workflow 系统研究

付鹤岗 武聪 尤娟

(重庆大学计算机学院 重庆 400044)

**摘要** 网格 workflow 可以定义成为完成特定目标而在异构和分布的资源上以一定次序执行的网格应用服务的集合。在网格 workflow 中引入服务质量,为网格服务的调度提供了依据,使得 workflow 的执行更加满足用户需求。通过使用服务质量可以对网格 workflow 进行选择 and 执行,从而更加符合用户的期望。提出了一种典型的基于服务质量的网格 workflow 管理体系结构及相关的服务质量参数体系,讨论了基于服务质量的工作流调度算法,并给出一个基于快速遗传算法的应用实例。

**关键词** 服务质量,网格,workflow,遗传算法

## Attribute-based Access Control Model for Web Services

FU He-gang WU Cong YOU Juan

(Computer College, Chongqing University, Chongqing 400030, China)

**Abstract** Grid workflow can be defined as the composition of grid application services which execute on heterogeneous and distributed resources in a well-defined order to accomplish a specific goal. Introducing of QoS can provide means for schedule of component services and make the execution of workflow satisfy the requirements of users better. QoS allows for the selection and execution of Grid workflow to better fulfill customer's expectations. We proposed an architecture for QoS-based grid workflow system and the QoS parameters of grid workflow were discussed. Some QoS-based schedule algorithms were discussed and a solution based on fast genetic algorithms was proposed.

**Keywords** Quality of service, Grid, Workflow, Genetic algorithm

## 1 引言

workflow 管理的主要功能是使业务处理过程自动化,协调人和各种应用程序工作。网格 workflow 是一种新型的工作流管理系统,旨在利用网格环境中丰富的、异构的和分布的资源,创建一种管理网格中的高性能计算能力的协同管理机制,实现在动态变化的多个虚拟机构间的资源共享和协同工作<sup>[1]</sup>。通过对已有 workflow 系统的研究发现,无论是成熟的产品还是原型系统,在提供一系列管理业务流程的重要功能的同时,却缺少了对服务质量的相应管理,而服务质量对于 workflow 系统来说也是一个很重要的问题。在网格 workflow 中引入服务质量,将应用所需的服务质量作为服务选择的依据,可以使基于 workflow 的组合服务的执行和调度围绕用户所提出的服务质量参数进行,使网格中的资源更好地围绕用户的要求进行组织和分配,workflow 的执行也更加满足用户的要求。

## 2 服务质量(QoS)

网格 workflow 支持的服务质量<sup>[2]</sup>取决于网格服务所支持的服务质量。本文从服务质量的重要性程度和易于度量的角度出发,对服务质量参数进行了选择,建立了包括如下 5 个服务质量参数的参数体系<sup>[3]</sup>。

(1)时间:用来评测服务完成请求的速度,是服务质量评估的一种常见和通用的度量指标。对于 workflow 系统来说,时间是指完成一次 workflow 执行或者响应一次服务请求所需要的总时间,可以进一步划分为请求等待时间和请求处理时间。

(2)成本:用来衡量 workflow 或者服务一次执行所需要的耗费,也可以进一步划分为管理和监控 workflow 系统的管理成本和任务实现成本。

(3)可靠性:指服务或者 workflow 在一定条件下,在特定的时间内执行所需功能的能力。可靠性可以用执行的成功率来衡量,也就是服务或者 workflow 执行成功的次数与总执行次数的比率。

(4)可提供性:可提供性与可靠性相关,指服务或者 workflow 在一定的条件下,在特定的时间内提供所需要的功能的能力。可提供性可以用服务或者 workflow 提供运行的时间与总时间的比率来衡量。

(5)逼真度:用于描述服务的可信度,可以反映出服务执行的好坏程度,它依赖于用户对于服务使用的体验。因为不同的用户可能对相同的服务有不同的体验,也就是说会给予服务不同的评价。逼真度可以采用用户使用某一服务后给予该服务的平均等级来衡量。

与传统的工作流系统相比,基于服务质量的工作流系统

到稿日期:2008-11-20 返修日期:2009-02-26

付鹤岗 副教授,硕士生导师,主要研究方向为软件工程、电子商务, E-mail: hgfu@cqu.edu.cn; 武聪 研究生,主要研究方向为电子商务、网格计算。

有以下特征:

(1) 基于服务质量的设计。方便用户更有效地将已有的商业设想转变为业务流程,因为 workflow 系统可以根据服务质量的评价标准来设计。

(2) 基于服务质量的选择和执行。根据相应的服务质量对 workflow 系统进行选择和执行,可以更好地实现用户的预期要求。通过对服务质量的分析,可以保证提供的服务满足用户需求。

(3) 服务质量的监控。必须在工作流系统整个服务生命周期内,对服务质量实行严格的监控,以保证初始的服务质量要求与目标服务质量相一致。无法满足服务质量的要求时,通过服务质量监控可以采用自适应策略做出相应的调整。

(4) 基于服务质量的调整。如果要实现一个满足初始服务质量要求的工作流系统,必须针对流程、延迟等技术因素,重新调整和规划 workflow 系统。而在运行 workflow 调整策略前,还需要明确其对 workflow 服务质量产生的影响。

### 3 系统架构

基于服务质量的工作流管理主要包括 workflow 规范、服务发现和工作流调度。如上所述,基于服务质量的工作流系统具有区别于传统 workflow 系统的特殊要求。为了满足 workflow 管理系统对于服务质量的要求,本文提出了一个典型的基于服务质量的工作流管理体系结构,如图 1 所示。workflow 管理系统的各部分如下。

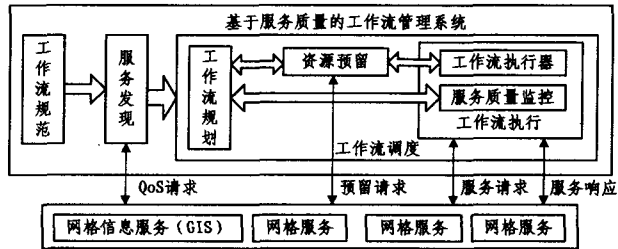


图 1 基于服务质量的工作流管理体系结构

#### 3.1 workflow 规范

基于服务质量的工作流管理允许用户通过使用 workflow 的规范来表示他们对服务质量的要求和任务描述以及相关的依赖关系。通常服务质量的约束条件可以表达用户的偏好,这对于资源的有效分配是必要的。

#### 3.2 服务发现

在提交了 workflow 规范后,workflow 系统就需要找到合适的服务来执行任务。在一个复杂的 workflow 中,不同的任务需要不同类型的服务来执行。而在一个服务网格中,即使对于同一类型的服务,它们也可能是由不同的服务提供者提供的,并且分布在多个管理域中。workflow 系统必须能够查询网格信息服务 (GIS),并生成一个与每个任务相对应的可用服务列表,方便 workflow 用户的使用。

在一个服务网格中,执行同一个任务的服务质量,其服务属性可能是不同的。因为不同的服务提供者可能会提供不同的服务质量,同一个服务提供者也可能提供不同级别的服务质量来满足不同用户的需求。

#### 3.3 workflow 调度

workflow 调度主要就是 workflow 任务映射到网格服务上,

并在网格服务上执行任务。对于服务网格来说,workflow 调度中的调度方案必须符合用户对服务质量的约束条件。workflow 调度主要有 3 个步骤: workflow 规划、资源预留、workflow 执行。

#### (1) workflow 规划

workflow 规划是在 workflow 中为每个任务选择一个使用的服务,并在 workflow 执行前生成一个调度列表。调度的结果应该满足用户对服务质量的相关约束条件。

对 workflow 执行的规划需要从服务质量的请求中获得服务质量的参数,然后对 workflow 进行整体的规划。将 workflow 任务映射到分布式的服务中通常是一个 NP 难题,因此 workflow 规划由于时间的约束,只能生成一个次优的调度规划。

#### (2) 资源预留<sup>[4]</sup>

资源预留这个概念最早出现在有服务质量保证的服务中<sup>[5]</sup>,它对于 workflow 调度特别是长周期的 workflow 执行极为重要。workflow 管理系统需要提前对已经被 workflow 规划选择使用的服务进行相关资源的预留,从而保证服务的可用性。

#### (3) workflow 执行

由于服务失效或服务延迟等原因,workflow 系统中的服务有可能会违反最初对服务质量的约定。因此,workflow 调度必须能够根据资源的动态变化进行更新调整。举例来说,如果一个任务的执行有延迟,在预期的时间内无法完成,就应该调度预留的资源来执行未完成的任务。在系统中通常使用服务质量监控器来监控服务性能的变化,并及时通告给 workflow 规划。

### 4 调度算法

#### 4.1 调度分析

网格 workflow 的调度过程就是控制 workflow 中各个成员服务的执行次序,并且从大量候选的网格服务中选择出合适的服务来执行 workflow 任务,其核心功能是对服务的选择。服务质量为网格 workflow 执行过程中服务的选择提供了一种依据。基于服务质量的网格 workflow 调度是基于网格信息服务 (GIS) 中提供的服务质量信息来进行的。所有网格服务的服务质量参数都注册到网格信息服务的信息表中,通过网格信息服务提供的查询接口,可以获得网格服务的服务质量信息。而 GIS 信息表中的内容也需要及时更新,才能够反映网格中服务的当前状况。随着 workflow 任务和服务数量的增加,搜索的空间急剧扩大。对于这类复杂的非线性组合优化问题,遗传算法是寻求满意解的最佳工具之一。

#### 4.2 遗传算法

遗传算法 (GA) 是由 John Holland 及其合作者在 20 世纪 70 年代提出的,为模仿自然界的自适应过程提供的一种搜索型算法,现在已成为求解调度问题在内的各类优化问题的有效优化算法之一。其操作的对象是“染色体”,它代表了所有优化问题解的一个二进制代码串,“染色体群”即为优化问题解的集合,而遗传算法搜索最优解的过程就是寻找适应度高的个体染色体。为此,将染色体群置于问题的环境中,基于适者生存的原则,通过对染色体的编码、种群初始化、选择、交叉、变异等过程实现染色体的进化。按此方式,通过迭代不断地进化,直至搜索过程收敛到最适应环境即适应度最高的一个染色体,从而求得问题的一个最优解。

#### 4.3 快速遗传算法

由于遗传算法通常以随机生成初始的种群(初始染色体群)为起点,因此最初得到的解的适应度往往比较低。经过多次的迭代之后,得到的解适应度逐渐提高,如图2所示。提高遗传算法搜索性能的关键是如何跨越适应度较低的初始阶段,从一个比较接近最优解的地方开始搜索<sup>[6]</sup>。要解决这一问题,需要充分利用已有的调度记录,因为提交到实际网格平台的任务之间通常都存在时间和空间上的关联关系<sup>[7]</sup>,也就是说先前提交的任务以后很可能再次出现。为此,在实际的调度中,本文采用快速遗传算法(FGA),保留并使用任务的调度记录来形成初始的染色体群。

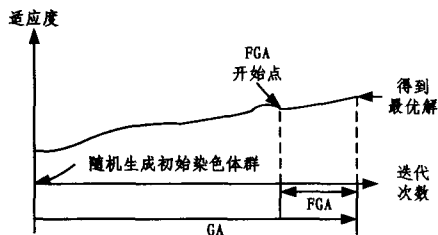


图2 快速遗传算法

## 5 实例分析

为了验证基于服务质量的网格 workflow 调度算法,本文选择使用 Gridbus<sup>[8]</sup>工具平台,通过实验来模拟大通关的工作流系统应用。其中,工作流的每个任务都需要相应的网格服务来执行。Gridbus 是由墨尔本大学提供的专门用于支持网格 workflow 开发的开源软件平台。大通关是涉及海关、外经贸主管部门、银行、保险等机构的系统,可以减少审批程序和办事环节,达到通关效率的大幅度提高。

大通关的主要流程如图3所示。在货物开始接受实际检查时,首先要进行申报的提交(t1)。申报表提交并接收后,海关开始对申报的受理(t2)。海关在受理过程中,还需要外经贸主管部门和银行等机构对申报进行审批(t3, t4, t5)。之后海关对货物进行预防性检查(t6)。检查完毕,就对货物放行(t7)。在检查过程中可能会出现查证挂起(t8),此时货物原则上可以被放行。

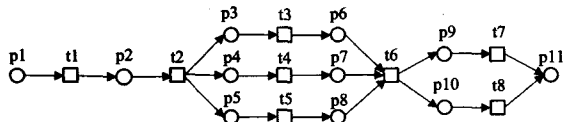


图3 大通关的 Petri 网模型

在实验中模拟了大通关 workflow 系统需要的8种网格服务,其中每种服务都可以由3个网格节点作为服务的提供者来执行。对于遗传算法的相关参数也进行了设定。初始染色体群数设为100,变异概率为0.04,而交叉概率是0.8。图4中显示了使用快速遗传算法(FGA)进行 workflow 调度的执行时间与迭代次数的关系。由图可知,算法的执行时间在前25次迭代中不断地波动变化,但是在第40次迭代时开始趋于收敛,并在第50次迭代之后逐渐达到一个平稳的值。实验数据有力地证明快速遗传算法对于基于服务质量的网格 workflow 调

度的有效性。

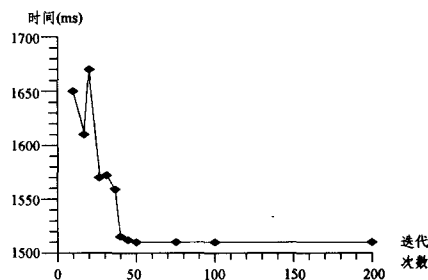


图4 快速遗传算法的完成时间

**结束语** 本文针对基于服务质量的网格 workflow 系统构建了系统评价服务质量的参数体系,详细介绍了典型的基于服务质量的网格 workflow 系统的体系结构。同时,对于网格 workflow 的调度问题,也提出一种基于服务质量的网格 workflow 调度算法。最后,使用一个 workflow 的应用实例对快速遗传算法进行了仿真实验和分析。为了简化实验,网格 workflow 中的每个服务只支持一种活动,而实际应用中的网格服务则要复杂得多。下一步的工作将针对更加复杂的网格 workflow 应用,研究网格 workflow 的动态调度问题。

## 参考文献

- [1] Chen J, Yang Y. Flexible Temporal Consistency for Fixed-time Constraint Verification in Grid Workflow Systems[J]. Lecture Notes in Computer Science, 2005, 3795: 300-311
- [2] Cardoso J, et al. Modeling Quality of Service for Workflows and Web Service Processes[J]. Web Semantics Journal: Science, Services and Agents on the World Wide Web, 2004, 1(3): 281-308
- [3] Patel C, et al. A QoS Oriented Framework for Adaptive Management of Web Service-based Workflows[J]. Lecture Notes in Computer Science, 2003, 2736: 826-835
- [4] Smith W, Taylor V. Scheduling with Advanced Reservations[C] // Proc. of the Int. Parallel and Distributed Processing Symposium Conf. Cancun, Mexico, 2000: 127-132
- [5] Benkner S, Brandic I. A Service-oriented Grid Environment for on Demand Supercomputing[C] // the Fifth IEEE/ACM International Workshop on Grid Computing, Pittsburgh, USA, 2004
- [6] Auyeung A. Multi Heuristic List Scheduling Genetic Algorithm for Task Scheduling[C] // Proc. ACM Sym. Applied Computing, 2003: 721-724
- [7] Kim S, Weissman J B. A Genetic Algorithm Based Approach for Scheduling Decomposable Data Grid Applications // Proc. of ICPP, 2004
- [8] Buyya R, Venugopal S. The Gridbus Toolkit for Service Oriented Grid and Utility Computing[C] // Proceedings of the First IEEE International Workshop on Grid Economics and Business Models. New Jersey, USA, 2004
- [9] 李秉智, 赵娜. 基于独立大粒度任务网格调度算法的分析比较[J]. 重庆工学院学报, 2007, 21(1): 57-61