

云工作流的调度算法

论文题目汇总

组 员	学号	论文题目
刘 硕	16340154	<i>Cost Minimized PSO based Workflow Scheduling Plan for Cloud Computing</i>
		<i>Scheduling Using Improved Genetic Algorithm in Cloud Computing for Independent Tasks</i>
刘 虹 奇	16340148	<i>Workflow Execution Plan Generation in the Cloud Computing Environment Based on an Improved List Scheduling Algorithm</i>
		<i>Scheduling scientific workflow tasks in cloud using swarm intelligence</i>
聂 博 业	16340171	<i>Workflow Scheduling Using Hybrid GA-PSO Algorithm in Cloud Computing</i>
许 博 阳	15331337	<i>A cost-effective deadline-constrained dynamic scheduling algorithm for scientific workflows in a cloud environment</i>

问题描述

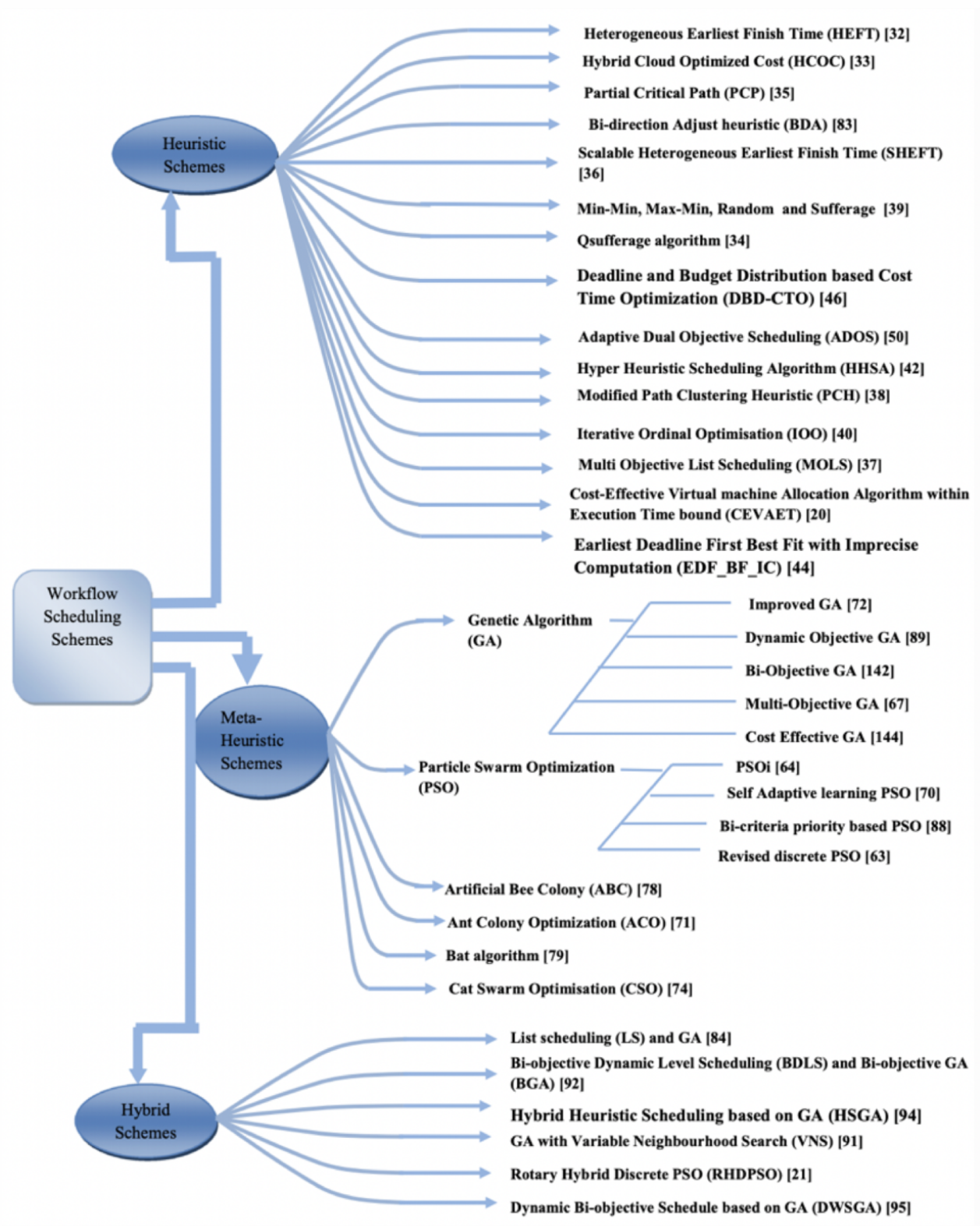
工作流调度考虑到各种服务质量需求，关注复杂任务到云资源的映射。随着云计算探索的不断扩展，在用户规范下寻找合适的工作流执行调度方案变得越来越紧迫。我们组重点研究了各种调度方案的综合分类，并对它们进行广泛的比较。另外我们还探索了这个场景下尚未解决的问题和可行的研究新方向。

现实背景

目前工作流调度 (WFS) 在分布式计算场景中十分流行，因为它能够使得相互依赖的任务映射到虚拟机 (VM)，完成在指定的服务质量 (QoS) 需求。在云计算中，IaaS、PaaS和SaaS等服务基本上是在服务水平协议 (SLA)，在QoS约束的帮助下由用户使用的。通常工作流是通过依赖关系绑定在一起的复杂任务流。工作流具有广泛的应用程序，可以在云和网格等分布式计算环境中编程。此外，工作流在截止日期、资源利用率等方面具有特定的约束，而在其他调度中，决策是为了命令执行彼此无关的独立任务。先前的研究已表明工作流上的任务可以转向云计算以达到高性能，我们小组主要关注了在使用分布式计算技术的情况下将WFS问题纳入云环境的研究。

研究现状

对于工作流的调度最近的工作如下：（来自 *Quality of Service (QoS) Aware Workflow Scheduling (WFS) in CloudComputing: A Systematic Review*。）



启发式算法

启发式的意思是“通过尝试和错误来发现”。因此这类算法包括能够在合理时间内提供优化问题解的算法，但不能保证达到最优解。当我们不一定想要最好的解决方案，而是想要容易达到的好的解决方案时这是很好的。

我们组对于启发类研究的论文是[Workflow Execution Plan Generation in the Cloud Computing Environment Based on an Improved List Scheduling Algorithm](#)，基于改进列表调度算法的云计算环境中的工作流执行计划生成。方法的具体介绍可以参考[这个链接](#)。我们还参考了[A cost-effective deadline-constrained dynamic scheduling algorithm for scientific workflows in a cloud environment](#)，研究了动态启发式算法在云科学工作流上的应用。方法的具体介绍可以参考[这个链接](#)。

元启发式算法

元的意思是“超越”或“更高层次”，一般来说，元算法的性能比简单的启发式更好，使用了一定的随机化和局部搜索的权衡。在这里，随机化为逃避局部搜索提供了很好的解决方案，因此所有的元启发式算法都适合全局优化。

我们组对于元启发类研究的论文是[Scheduling scientific workflow tasks in cloud using swarm intelligence](#)（ACO），利用群智能调度云中的科学工作流任务，方法的具体介绍可以参考[这个链接](#)。另一个关于这个方面的研究是一个基于BPSO的改进方案，是原作者在之前工作上的扩展，[Cost Minimized PSO based Workflow Scheduling Plan for Cloud Computing](#)。方法的具体介绍可以参考[这个链接](#)。

融合型算法

这一类别包括启发式和元启发式算法的混合，也包括元启发类算法之间的hybridization。

我们组对于混合类研究的论文首先参考了[Scheduling Using Improved Genetic Algorithm in Cloud Computing for Independent Tasks](#)，提出了一种改进的GA算法，融合了min-min和min-max，方法的具体介绍可以参考[这个链接](#)。另一种融合算法的论文是[Workflow Scheduling Using Hybrid GA-PSO Algorithm in Cloud Computing](#)，这个论文也是把常见的基因算法和粒子群智（PSO）进行了融合，方法的具体介绍可以参考[这个链接](#)。

结论和展望

云工作流中相关论文研究的方法主要有启发类、元启发类和混合方案。总体而言，启发类算法的研究时间普遍较早集中在21世纪第一个十年（当然也有后期新的进展），有着更强创新性的元启发类应用了群智算法、基因算法等新颖的技术。最近几年大部分都是一些混合方案，都是对之前方法的改良，原创性不强。

关于云环境下工作流的调度问题仅仅是云 workflow 问题分支的一个，在工作流系统的各种难题中也只是沧海一粟。即使如此，我们仍未对这个问题拥有足够深入的认知。在以往的研究中，完工期和成本是迄今为止研究人员关注的唯二目标。未来这项研究仍然拥有广阔的空间：

- 可以考虑一下云资源的容错性（近两年只查到两篇 [Fault-Tolerant Scheduling for Scientific Workflow with Task Replication Method in Cloud](#)、[An Efficient Fault Tolerant Workflow Scheduling Approach using Replication Heuristics and Checkpointing in the Cloud](#)）、负载均衡（近两年只有两篇 [A Hybrid Meta-heuristic Approach for Load Balanced Workflow Scheduling in IaaS Cloud](#)、[A PSO-based task scheduling algorithm improved using a load-balancing technique for the cloud computing environment](#)）和工作流的安全性（近两年只查到一篇 [Genetic-based algorithms applied to a workflow scheduling algorithm with security and deadline constraints in clouds](#)）。
- 可以思考工作流失败的灾难性后果并通过机器学习等方法预测失败，从而减少工作流失败对云资源的灾难性后果。（近两年甚至还没有！！）
- 当然也可以继续实现新的启发式算法和元启发式算法以及融合型。

附录

名词解释

- 简单工作流：定义了简单作业的执行，其中的任务集以DAG的形式表示。
- 科学工作流：包含了大量复杂的数据，以数千个任务的形式表示为DAG。
- 分布式计算：随着计算技术的发展，有些应用需要非常巨大的计算能力才能完成，如果采用集中式计算，需要耗费相当长的时间来完成。分布式计算将该应用分解成许多小的部分，分配给多台计算机进行处理。这样可以节约整体计算时间，大大提高计算效率。
- 网格计算：网格计算研究如何把一个需要非常巨大的计算能力才能解决的问题分成许多小的部分，然后把这些部分分配给许多计算机进行处理，最后把这些计算结果综合起来得到最终结果。
- 调度：具有指定用户约束时，在资源上处理任务映射的行动步骤。
- NP问题：在计算机学科中，存在多项式时间的算法的一类问题，称之为P类问题；而像梵塔问题、推销员旅行问题、可满足问题这类，至今没有找到多项式时间算法解的一类问题，称之为NP类问题。而NP问题中最困难的问题称之为NP完全问题（NP-complete）。
- 遗传算法：遗传算法是模拟达尔文生物进化论的自然选择和遗传学机理的生物进化过程的计算模型，是一种通过模拟自然进化过程搜索最优解的方法。
- 蚁群算法：是一种用来寻找优化路径的概率型算法。

其他参考

论文主纲参考了论文 *Quality of Service (QoS) Aware Workflow Scheduling (WFS) in Cloud Computing: A Systematic Review*。

论文查阅网站包括 **Research Gate**、**Arxiv**、**Springer**、**IEEE Explore**。大部分阅读论文均为近三年的SCI和EI论文。
