

关于动态资源分配的相关研究

邓振瀚

华中科技大学 土木与水利工程学院

摘要 资源分配是计算机系统中的一个基本问题，但现今对资源最热门的分配方法都是基于静态进行的研究，对于动态的用户需求，很明显静态的资源分配不再满足帕累托效率、策略验证性和公平性，但实际生产过程中，用户的需求是动态变化的，这就需要研究适应动态需求的资源分配机制。本文对众多关于动态资源分配的研究进行了相关解读，希望从中获取到现如今关于动态资源分配的合理解决方法与相关研究机制，并希望能对今后的研究起到一定的指导作用。

关键词 资源分配；帕累托效率；策略验证性；公平性动态；动态分配

Related research on dynamic resource allocation

Hanzhen Deng

HUAZHONG UNIVERSITY OF SCIENCE AND TECHNOLOGY School of Civil and Hydraulic Engineering

Abstract Resource allocation is a basic problem in computer systems, but the most popular resource allocation methods are based on static research. For dynamic user needs, it is obvious that static resource allocation can no longer meet the Pareto efficiency, policy validation and fairness, but in the actual production process, user needs are dynamically changing. So it is necessary to study the resource allocation mechanism that ADAPTS to dynamic demand. This paper interprets many researches on dynamic resource allocation, hoping to obtain reasonable solutions and relevant research mechanisms on dynamic resource allocation, and hopes to play a certain guiding role for future research.

Key words * Resource allocation; Pareto efficiency; Strategy-proofness; Fairness; Dynamic allocation

1 引言

资源分配是计算机系统中的一个基本问题，涉及私有云和公共云、计算机网络、管理程序等。在设计资源分配机制方面，有大量活跃的研究，以实现帕累托效率(高资源利用率)和策略验证性(自私的用户不应该通过谎报需求而受益)，同时确保资源在用户之间公平分配。

我们考虑在用户需求是动态的系统中公平分配资源的问题，也就是说，用户需求随时间变化。我们的主要观察是，用于资源分配的经典最大最小公平算法提供了许多理想的属性(例如，帕累托效率、策略验证性和公平性)，但只有在用户需求随时间保持静态的强烈假设下。对于动态用户需求的现实情况，最大最小公平性算法失去了这些属性中的一个或多个。

2 背景

现两种最流行的分配机制是严格分区和最大最小公平。严格分区将资源平均分配给所有用户(“公平分享”)，与用户的需求无关；这保证了策略验证性和公平性，但不能保证帕累托效率。最大最小公平通过考虑用户需求来减轻严格分区的限制：它最大化用户之间的最小分配，同时确保每个用户的分配不超过他们的需求，基于最大最小公平的分配保证了三个理想属性——帕累托效率、策略验证性和公平性。以上方法均为基于静态提出的分配机制，但实际生产过程中，用户的需求是动态变化的，这就需要研究适应动态需求的资源分配机制。

3 研究现状

R. Anantha Kumar 等人研究了利用动态定价拍

卖机制支持云计算应急需求的资源分配。云资源提供了各种类型的(VM)虚拟机请求,这些请求在一个精确的时间范围内分配给客户端。目前,在 Cloud 环境中,虚拟机的调度方法是通过固定价格调度算法来决定的,用户在单位时间内支付固定的费用来获得资源。然而,这种调度算法在云上并不有效,即使云资源是动态分配和释放的。为了解决这一问题,他们提出了一种基于动态定价的组合拍卖分配机制的自适应调度算法。通过组合拍卖的动态定价,提高资源利用率和用户满意度。他们提出的基于市场的调度算法利用拍卖机制的原理,以扩大云供应商和客户的满意度。该技术重构了当前资源分配的偏好,以便在虚拟机出现紧急需求时提前分配资源。然后论证了集体目标增强数值原型,形成了连接物理机与虚拟机的最小执行等效范围,资源分配的目标是获得最少数量的物理机。仿真实验结果表明,所提出的调度方法和集体目标增强数值原型能够充分提高服务质量(QoS),提高供应商利润和资源利用率。

本文提出了动态定价策略和动态用户需求,以增强云环境下资源供应商的竞争力,提高资源利用率。提出了一种新的基于组合拍卖的动态定价分配机制来分配资源以获得云流。利用所提出的技术,资源供应商可以通过增加额外收入来转移成本以提高拍卖资源的机会,从而提高资源的利用率。客户选择最优资源,包括最小组合拍卖 CATC (Product of Time and Cost),以更低的资金成本保证更短的终止时间。最后,对所提出的估计方法进行了评估,包括典型的 DPAM 过程。结果表明,该技术可以实现更高的资源利用率,更短的终止时间,更低的资金成本。利用动态定价机制,供应商可以降低个体资源成本,增强竞争能力。对于意外和紧急情况,云资源的提供应该适当和准确。本文提出了一种适应紧急用户需求的云资源动态分配技术。该机制提供了资源分配的优先级,以保证紧急资源请求的合理分配,并建立了一个集体目标增强数值原型。这利用了物理系统的最小量,而虚拟系统和物理系统之间的相似性最大。

未来,扩大成本将与动态定价系统相关联。对于个体资源供应者来说,这仍然是一个最好的方法,他们通过更大的竞争来扩大成本以获得额外收入。同时,它可以应用于标准的系统数据集来执行试验和不规则信息。它将扩大审判结果的完整性,并为模拟 DPCAM 方法的执行提供额外的系统。此

外,可以通过包括进一步的服务质量来检查带有资本成本的终止时间。

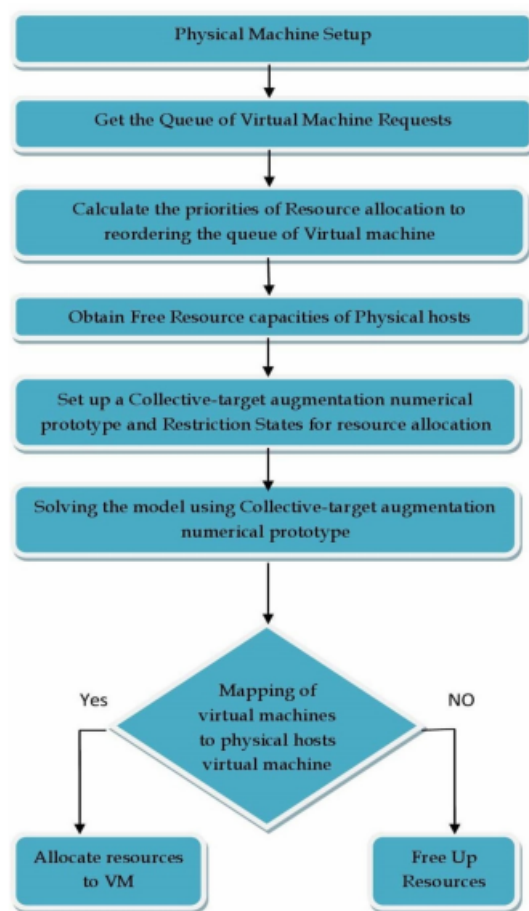


Fig. 1. Resource allocation method flow graph.

图 1 R. Anantha Kumar 等人提出的资源分配方法流程图

ALI MOAZENI 等人研究了云环境下基于自适应多目标教学优化算法的动态资源分配。资源分配是云数据中心中的一个非多项式完备问题,它选择合适的资源来执行许多精细的计算粒度任务。客户需求和应用程序的容量经常变化。为了弥合频繁变化的客户需求与服务可用基础设施之间的差距,他们提出了一种使用云计算中自适应多目标基于教学的优化(AMO-TLBO)算法的动态资源分配策略。AMO-TLBO 引入了教师数量、适应性教学因子、导师制培训和自主学习的概念,以提高教师的探索和开发能力。此外,还使用基于网格的方法自适应地评估外部存档中维护的非主导解决方案。AMO-TLBO 的目标包括最小化完工时间、成本和最大化利用跨虚拟机的均衡负载。评估结果表明,该算法在不同性能指标上均优于 TLBO、MOPSO 和 NSGA-II 算法。

在云计算中,动态资源分配是一个优化调度的

过程，它增加了负载均衡的数量。云资源分配对云服务提供商和云服务客户都非常重要。客户倾向于在较短的时间内以较低的成本处理请求，而利润最大化是云提供商最重要的优先事项之一。因此，必须同时满足多个相互冲突的目标。提出了一种基于自适应多目标教与学的云计算资源分配优化算法。该算法在缩短完工时间、降低总成本和提高资源利用率等目标之间做出了妥协。此外，在本文提出的算法中，在 TLBO 中使用了一个自适应教学因子，该因子可以根据学习者和教师的结果在搜索过程中自适应变化，以提高算法的性能。此外，采用基于网格的方法自适应地评估保存在外部存档中的非主导解决方案。通过与 TLBO、MOPSO 和 NSGA-II 算法的比较，证明了该算法的有效性。仿真结果表明，该算法能显著降低用户的最大完工时间和总成本，提高资源利用率。

在未来的研究中，将评估其他技术，以提高 AMO-TLBO 的能力，并考虑其他目标，如能源效率，使用数据挖掘技术，如发现重复模式，以帮助解决资源分配问题。

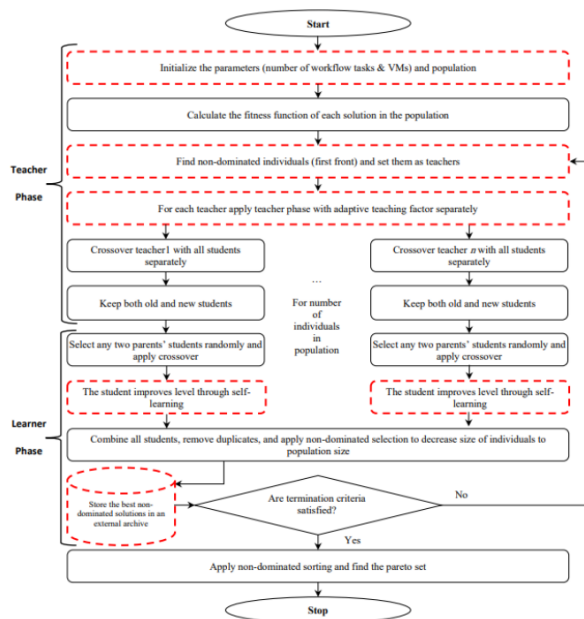


图2 ALI MOAZENI 等人提出的算法流程图

Sakshi Chhabra 等人研究了云数据中心网络负载均衡调度的动态资源分配方法。云数据中心有许多主机和应用程序请求，其中资源是动态的。对资源分配的要求是多种多样的。这些因素可能导致负载不平衡，从而影响调度效率和资源利用率。提出了一种基于负载均衡的动态资源分配(DRALB)调度方法。证明了 SLA 感知调度不仅通过资源可用性

促进云消费者，提高吞吐量、响应时间等，而且在减少资源利用率和 SLA(服务水平协议)违反处罚的情况下最大化云利润。该方法基于客户端应用程序的多样性，为特定的部署搜索最优资源。实验基于以下参数:平均响应时间;资源利用率、SLA 违规率和负载均衡。实验结果表明，该方法可以减少资源的浪费，减少网络中高达 44.89%和 58.49%的流量。

本文提出了一种基于应用 SLA 的资源分配方案，用于分析资源需求并合理分配资源数量用于特定部署的物理机器。该框架监视即将到来的应用程序，并演示支持 SLA 的调度策略。这两个性能参数的重点是在监视应用程序时满足每个 SLA 违反的公平性约束。实验表明，与顺序、随机和 DHLB 方法相比，DRALB 方法可以提高有效的负载均衡。这种支持 SLA 的云负载均衡器最大限度地提高了云服务的优势，减少了故障数量，并将惩罚函数降至最低。性能评估表明，DRALB 在指出现有知名算法的观察结果的同时，取得了令人满意的输出，减少了实验数据的资源浪费，减少了流量，分别达到 44.89%和 58.49%。

在未来的工作中，将研究自适应方法，以更好地平衡 SLA 违反与资源效率之间的权衡。此外，在模型中加入更多的目标，然后在真实的云数据中心中实现算法。

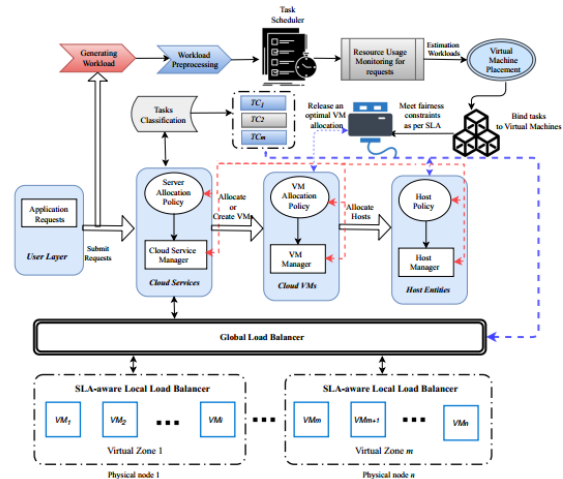


图3 DRALB 框架

Xiaotian Pan 等人研究了协同约束多目标优化算法的动态资源分配策略。不可行解有助于找到可行区域，但需要投入多少可行解和多少不可行解才能达到最优的搜索效率还有待进一步研究。结合最近提出的协作约束多目标框架，讨论了不同类型 CMOPs 中辅助群体和原始群体的贡献。在不同的 CMOPs 和不同的搜索阶段对这两个种群分配相同

的资源是不合理的。本文旨在研究约束环境下的资源分配，以有效地利用有限的资源，获得更好的性能。因此，首先引入投资回报率(ROI)的概念来衡量两个种群的贡献，然后指导种群规模分配(APS)。为了防止投资回报率随着种群规模的减小而不断下降，提出了一种进化资源分配策略(AER)，根据合作关系调整它们的进化状态，进一步提高它们的投资回报率，并再次竞争种群规模，使两个种群在竞争和合作中的进化效率最大化。将所提出的 CCMODRA 与涵盖三种 CMOEAs 的七种流行算法进行比较，并在涵盖四种 CMOPs 的三个基准上进行测试。CCMODRA 在三目标 CDRLZs 上的综合性能优于其他七种 CMOEAs，在五目标 CDRLZs 上的综合性能优于 71%，在五目标 CDRLZs 上优于 57%，在 MWs 上优于 46%。通过生成贡献解和 DOC 测试问题，验证了 APS 和 AER 策略的有效性。此外，CCMODRA 算法在有容量约束的背包问题上的总利润比其他七种算法提高了 0.2%~216%。

本文提出了一种基于约束的多目标协同优化的动态资源分配策略。对于四种类型的 CMOPs，详细分析了可行解和不可行解对整个演化过程的贡献差异。为了充分利用两个种群之间的合作关系来探索受约束的目标空间，首先引入 ROI 的概念来计算两个种群对进化的贡献差，然后对这两个种群进行资源的动态分配，包括种群规模分配(APS)和进化资源分配(AER)。本文提出的 CCMODRA 在四种类型的 CMOEAs 上进行了测试，在 50% 以上的 3-objective 和 5-objective C-DTLZs 上显示出更好的综合性能，46% 以上的 MWs 优于其他 7 种 CMOEAs。采用 APS 和 AER 策略可以显著提高 CCMO 生成贡献解和求解决策约束 CMOPs 的能力。此外，在 MOKP 上验证了 CCMODRA 在实际 CMOPs 上的性能

在未来，有两个主题需要进一步探讨。其中一个主题是通过合作框架将 DRA 战略应用于更多的 CMOEAs。二是设计 DRA 策略，以解决 CMOPs 更复杂的特征，如生成更多具有更好多样性的解来逼近不连续的 PF，或在不可行的区域很小时进行跨区域搜索，在原始种群和辅助种群之间共享有用信息。

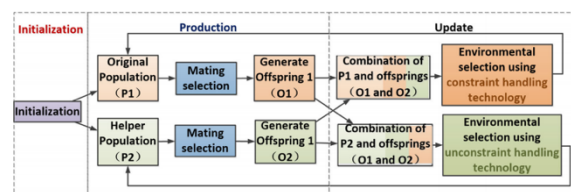


图 4 CCMO 的框架

NIAL CREECH 等人研究了动态多智能体系统中的资源分配。资源分配和任务优先级是自动驾驶汽车、网络和云计算领域的关键问题。开发高效和鲁棒算法的挑战来自于这些系统的动态性，其中许多组件以复杂的方式进行通信和交互。他们提出的多组资源分配优化(MG-RAO)算法使用随时间变化的资源需求的多个函数近似，以及强化学习技术，开发了一种优化这些多智能体系统资源分配的新方法。这种方法适用于对共享资源有竞争需求的情况，或者任务优先级问题。在模拟环境中，MG-RAO 算法比固定资源分配算法提高了 23~28%。结果还表明，在一个易变的系统中，使用配置为子代理对所有代理的资源分配建模的 MG-RAO 算法，其性能是设置为多组代理建模时的 46.5%。这些结果表明，该算法能够解决多智能体系统中的资源分配问题，并在动态环境中表现良好。这项工作研究了代理如何在多个任务之间分配资源的问题，这些任务已经作为其他代理的复合任务的一部分分配给它们，以提高系统的效用。提出的解决方案稿件提交给 ACM，使用 MG-RAO 算法根据分配任务的父代理将代理的任务分组在一起，随着时间的推移为父代理建模这些任务的值，然后将这些模型组合起来，以提供将提高系统总体效用的资源分配。评估结果表明，该方案在模拟系统中的性能比统一分配策略好 23~28%。由于子智能体学习了分配给它的任务对父智能体的价值模型，它能够更好地分配资源以最大化效用。在父代理组中划分入站任务的多个模型的使用允许学习比仅使用一个组时更详细的资源分配模型，并且还使算法在任务分布的波动性和变化下具有鲁棒性。在从 10-50 个父代理分配任务的 1-3 个子代理的系统中进行评估，代表了在 V2X 等现实系统中发现的常见范围。父代理组的添加及其可伸缩性能对于算法在大范围系统大小中的适用性非常重要。由于它们将父代理聚合到固定数量的组中，因此它们约束了运行算法所需的资源开销。这有助于解决方案适用于更大的系统。

未来的工作将着眼于该算法在现实环境中的

应用，如无线传感器网络中的资源分配、云计算环境、自适应服务组合以及 V2X 系统。此外，算法本身也有改进的地方。

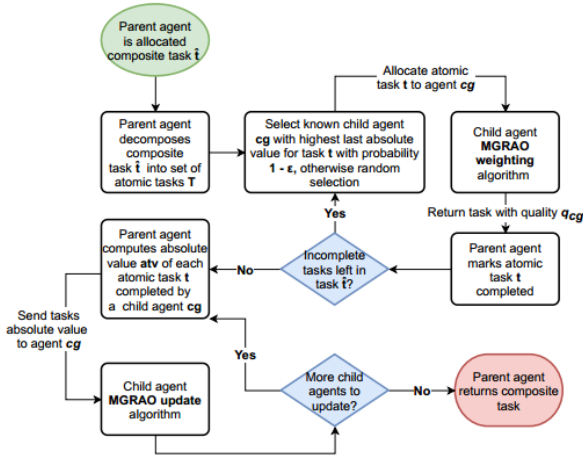


图 5 MG-RAO 工作流的高级视图

4 总结

在资源分配和调度方面有大量积极的工作，本文对众多关于动态资源分配的研究进行了相关解读，希望从中获取到现如今关于动态资源分配的合理解决方法与相关研究机制，并希望能对今后的研究起到一定的指导作用。效率、策略验证性和公平性

参考文献

- [1] Vuppapapati M, Fikioris G, Agarwal R, et al. Karma: Resource Allocation for Dynamic Demands[J]. arXiv preprint arXiv:2305.17222, 2023.
- [2] Kumar R A, Kartheeban K. Resource allocation using Dynamic Pricing Auction Mechanism for supporting emergency demands in Cloud Computing[J]. Journal of Parallel and Distributed Computing, 2021, 158: 213-226.
- [3] Moazeni A, Khorsand R, Ramezanpour M. Dynamic Resource Allocation Using an Adaptive Multi-Objective Teaching-Learning Based Optimization Algorithm in Cloud[J]. IEEE Access, 2023, 11: 23407-23419.
- [4] Chhabra S, Singh A K. Dynamic resource allocation method for load balance scheduling over cloud data center networks[J]. Journal of Web Engineering, 2021, 20(8): 2269-2284.
- [5] Pan X, Wang L, Zhang M, et al. A dynamic resource allocation strategy for collaborative constrained multi-objective optimization algorithm[J]. Applied Intelligence, 2023, 53(9): 10176-10201.
- [6] Creech N, Pacheco N C, Miles S. Resource allocation in dynamic multiagent systems[J]. arXiv preprint arXiv:2102.08317, 2021.
