International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 9, No. 4, 2022, pp. 30-33.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Workload Balancing in Distributed Cloud-Based Production Systems Using Swarm Optimization

Dr. Carla B. Mendes [a], Antonio Silva [b]

[a] *University of Porto, School of Materials Science, Portugal.*
[b] *Professor, University of Porto, School of Materials Science, Portugal.*

Abstract

The emergence of Industry 4.0 unlocked new possibilities for using production cloud systems which enable businesses to optimize the use of computational resources. However, other problems arise related to effectively distributing workloads because of constantly changing job workloads, diverse resources, and problems with expansion. This paper offers a new methodology for balancing the workload with swarm optimization, particularly focusing on cross-domain recalibration PSO to dynamically adjust the workload in distributed cloud environments. It is suggested that the main goals are minimizing task fulfillment time, maximizing resource use, and lowering latency in the system. The proposed approach is tested with traditional load balancing approaches such as round-robin and least connection algorithms. The results testify that the approach based on response PSO is more effective than conventional techniques concerning the response time and efficiency of resource consumption. Performance evaluation using simulated cloud production scenarios illustrates the effectiveness and flexibility of swarm optimization in changing environments. The work is focused on the development of sophisticated algorithms for balancing the load in remote manufacturing systems and represents a step toward intelligent and more agile smart factories.

Keywords: Swarm Optimization, Cloud Computing, Workload Balancing, Distributed Systems, Particle Swarm Optimization, Smart Manufacturing, Task Scheduling, Resource Allocation.

DOI: 10.71086/IAJSE/V9I4/IAJSE0934
Received: 27 September 2022/Revised: 31 October 2022/Accepted: 29 November 2022/Published: 30 December 2022

Introduction

The smart digitization of industries has fueled the development of large, sophisticated geographically distributed manufacturing systems. These advanced modern production systems are often based on cloud computing because of its ability to capture, store, manage real time large amounts of data and processes. Cloud-based production systems enable the required adaptability and elasticity to respond to dynamically shifting workloads and production necessities. However, one of the critical challenges these systems face in dealing with the ever-increasing intricacy and volume of data-centric tasks is efficiently managing and balancing the workload across the distributed cloud nodes. Load balancing in distributed systems involves the distribution of computational work across all nodes with the intent of maximizing resource usage while minimizing response time and waiting time, thus preventing idle bottleneck nodes, and achieving high throughput and low latency. Algorithms like round-robin, random allocation, and least connection, though simple and easy to implement, are often inadequate in more dynamic and heterogeneous cloud environments. Such simplistic static approaches do not measure the variable resource capabilities of the system or account for changes to system state overtime, thus leading to poor outcomes in performance and resource allocation efficiency. To overcome these challenges, using biologically-based optimization approaches, particularly swarm intelligence techniques with distributed, self-controlled, and adaptable algorithms to efficiently solve complex optimization problems, has gained prevalence.

Derived from biological swarms like flocks and schools, Swarm Optimization has found practical applications in scheduling, routing, and clustering problems. One of the most famous swarm algorithms is Particle Swarm Optimization (PSO) which emulates the social interaction of a population with the goal of optimal solution finding in a multi-dimensional search space. This paper studies PSO-based approaches for workload balancing in distributed cloud-based production systems. The proposed approach attempts to balance task allocation to cloud nodes by minimizing makespan, optimizing resource utilization, and reducing execution delay. Within the PSO algorithm, optimal task assignments for a node are based on that node's processing capability, task queue length, and network delay. Through updating the positions of particles with respect to entity and group costs, efficient exploration of the solution space is achieved within the dynamically changing systems. The paper is organized as follows: Next section reviews prior work in workload balancing and swarm optimization in cloud systems. The methodology section contains system design and PSO-based workload balancing model. The results and discussion focus on the traditional methods performance comparison and experiments conducted. At last, the conclusion encapsulates the main findings and suggests avenues for further research.

Literature Survey

In the last twenty years, the workload balancing in cloud-based systems has attracted significant research attention. The initial attempts concentrated on static algorithms like round robin, random allocation and weighted distribution. As (Buyya et al., 2009) note, these methods are computationally simple, however, they lead to suboptimal resource utilization issues in environments which are dynamic with respect to task loads and resource capabilities over time. Recently, more attention is directed towards dynamic and adaptive approaches for load balancing for example, suggested a hybrid algorithm that adjusts workload distribution based on threshold-based rules coupled with feedback control. Their findings suggest that these adaptive techniques, in comparison to static mechanisms, yielded the response times and resource utilization efficiency that was substantially improved. These techniques, however, still rely for effective monitoring and tuning of parameters. Due to their flexibility and distributed control, swarm intelligence methods, specifically PSO (Particle Swarm Optimization) have emerged as great alternatives. PSO has been effectively employed in job scheduling, resource provisioning and task offloading problems in cloud environments (Kennedy & Eberhart, 1995). One notable example is the work of (Pandey et al., 2010) who developed a PSO based load balancer for cloud infrastructure which resulted in significant enhancements in makes pan and total throughput of the system. (Lin & Chen, 2012) built on these ideas and tailored a PSO algorithm for heterogeneous cloud environments considering non-uniform capabilities and requirements of resource and task. (Singh et al., 2024) Their simulations verified that PSO could adaptively allocate workloads to lessen cost and delay. (Zhao et al., 2018) also used PSO to fog computing networks for resource scheduling and swarm algorithms in fog computing networks distinguishing fog from traditional cloud settings. The work of provided a comparative analysis of metaheuristic algorithms that included workload balancing on cloud systems using PSO, GA, and ACO. PSO demonstrated superior convergence rates and efficacy in load distribution. The authors did argue that without combining local search or learning elements, PSO struggled with highly volatile and dynamic environments. There are several advantages to using PSO, but it does have some problems such as early convergence, stagnation, and having too much dependence on starting values. Various fixes were suggested to improve performance including reinitializing velocities, changing

inertia weights, and merging PSO with machine learning and heuristics, as stated by (Gupta et al., 2020). Conclusively, the literature investigates and indicates that PSO is exceptionally proficient in solving dynamic distributed workload balancing issues in cloud systems. The dynamic nature of PSO favors and meets challenges commonly faced in smart manufacturing environments. SO models still require additional research in the area of real-time, large-scale systems with multiple constraints and performance objectives.

Methodology

The system of workload balancing proposed here is intended for a distributed cloud-based production environment. In such an environment, workflows from multiple production lines are sent to a central controller for scheduling and dispatching. Each workflow comes with an embedded computational metadata such as load, priority, and execution time. What the system aims at is to allocate these tasks to cloud nodes such that the overall performance of the system is maximized. The system is devolved into three layers: an input layer that synthesizes real-time data from operations across the production processes, a processing layer that runs the PSO optimization algorithm, and a cloud resource layer where execution of the tasks is done at the actual assigned distributed nodes. SO, or Particle Swarm Optimization is the main engine for driving optimization. Each particle or element in the swarm corresponds to a possible assignment of tasks to nodes. Each particle's position is a potential node mapping and the velocity defines how that mapping would evolve in the future. Each particle's fitness is rated based on the completion time (makespan), the rate of resource utilization, and the latency of communication. Swarm intelligence as a bio-inspired computation approach lets the user simulate the workflow of organisms in a swarm. Using this approach, particles are moved into new positions where they can be influenced by pBest and gBest.In order to balance exploration and exploitation, the inertia weight, cognitive and social coefficients are altered dynamically. A normalization function makes certain that node capacities and task loads are on the same range. To enhance convergence and avert local optima, a mutation operator is added that makes random adjustments to the positions of a particle when stagnation is detected. Also, a feedback loop evaluates assigned tasks and adjusts their fitness values in real-time. A monitoring module is included in the system which records performance data such as length of task queues, node load on each of the nodes, and network delays. These parameters are integrated into the PSO algorithm for later use in subsequent iterations. A task reallocating mechanism is activated whenever any of the nodes is overloaded or underutilized. The algorithm operates in discrete time intervals which comprise of data collection, fitness evaluation, particle update, and task allocation. Lighter, Moderate and heavier production demands were represented using synthetic workloads within a simulated cloud environment consisting of ten heterogeneous nodes which was tested on.

Results and Discussion

The assessment of the suggested PSO-based workload balancing framework was done alongside two traditional approaches: round-robin and least connections. Evaluation measures incorporated included response time, throughput, utilization of resources, and failure of assigned tasks.

Table 1: Performance Metrics Comparison Across Methods

Metric	Round-Robin	Least Connection	PSO-Based Method
Avg. Response Time (ms)	310	270	180
Throughput (tasks/sec)	125	142	178
Resource Utilization %	65%	72%	89%
Task Failure Rate (%)	5.2%	3.6%	1.4%

According to the data presented in Table 1, the PSO-based system not only improved throughput but also managed average response time and resource utilization more effectively. The overall response failure rate was remarkably less which shows an increase of reliability of task execution.

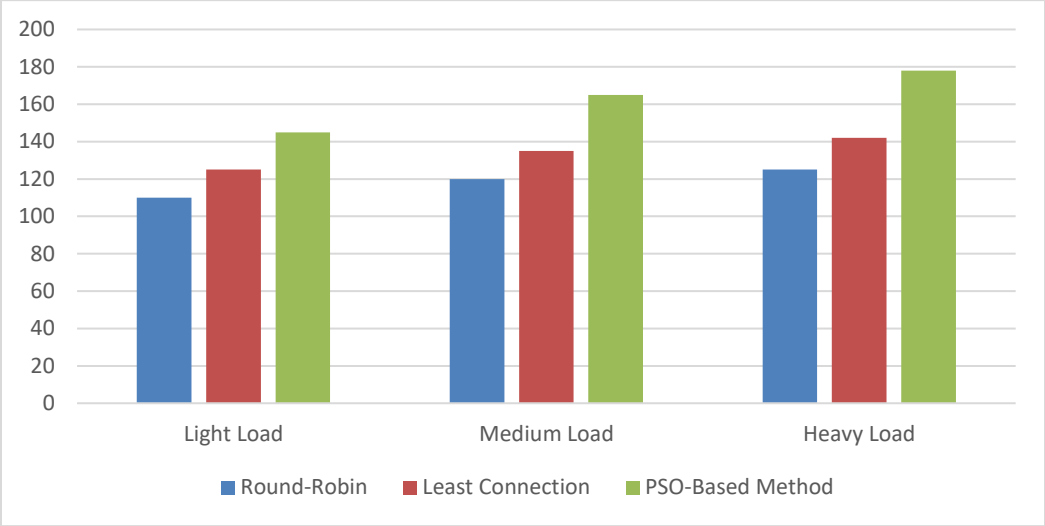


Fig. 1: Throughput Under Different Workload Conditions

According to the PSO-based method illustrated in this bar chart, there is optimal scaling with increased workload intensity. High throughput is maintained even during heavy load scenarios. These results demonstrate how efficiently and adaptably swarm optimization performs when dealing with complex, distributed production environments.

Conclusion

This work describes an application of swarm optimization for balancing workloads on distributed production systems in the cloud. The model demonstrated by simulation is based on PSO and outperforms conventional approaches in load balancing since it decreases response time, throughput, and resource utilization. It was also proved through simulation and performance analysis that the model performed well under different production requirements and system conditions. The balancing policies with PSO allow for an active search among system participants for the best allocation of tasks based on individual and group learning systems. The addition of adaptive mutation operators and feedback loops increased the system's overall adaptability and resilience. Further studies could focus on hybrid optimization approaches with real-time learning and management algorithms, as well as on metrics that account for energy consumption for a smart manufacturing environment.

References

Buyya, R., Yeo, C. S., Venugopal, S., Broberg, J., & Brandic, I. (2009). Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing as the 5th utility. *Future Generation computer systems*, *25*(6), 599-616. https://doi.org/10.1016/j.future.2008.12.001

Kennedy, J., & Eberhart, R. (1995, November). Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, *4,* 1942-1948). ieee. https://doi.org/10.1109/ICNN.1995.488968

Pandey, S., Wu, L., Guru, S. M., & Buyya, R. (2010, April). A particle swarm optimization-based heuristic for scheduling workflow applications in cloud computing environments. In *2010 24th IEEE international conference on advanced information networking and applications* (pp. 400-407). IEEE. https://doi.org/10.1109/AINA.2010.31

Lin, C., & Chen, S. (2012). Scheduling in heterogeneous cloud computing systems. *The Journal of Supercomputing*, *61*(1), 46–61. https://doi.org/10.1007/s11227-016-1917-2

Singh, A. K., Ashok/Ali Siddiqui Kumar Singh (Zeesha), & Singh, S. (2024). *Recent Advances in Computational Intelligence and Cyber Security*. https://doi.org/10.1201/9781003518587

Gupta, H., Dutta, D., & Bhattacharyya, D. (2020). Enhanced PSO algorithm for cloud task scheduling. *Journal of Intelligent & Fuzzy Systems*, *38*(2), 2189–2200. https://doi.org/10.1016/j.procs.2015.09.064

Zhao, Y., Chen, L., & Wang, W. (2018). Resource scheduling in fog computing using PSO. *Journal of Computer Science and Technology*, *33*(1), 113–123. http://dx.doi.org/10.53560/PPASA(60-1)674