Ijfans International Journal of Food and Nutritional Sciences
ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 10, Issue 01, 2021

A MULTI-OBJECTIVE ARTIFICIAL INTELLIGENCE FRAMEWORK FOR ENERGY-AWARE VIRTUAL MACHINE SCHEDULING IN CLOUD DATA CENTERS

Dr. Diwakar Ramanuj Tripathi

Assistant Manager I.T, Dinshaws Dairy Foods Pvt. Ltd., Nagpur

Email : drtcomptech@live.com

Abstract

However, cloud data centers require a lot of power consumption, and how virtual machines (VMs) are scheduled in the physical machines (PM) has a huge impact on the power consumption, resource utilization, and the quality of service provided to the users. In this study, an artificial intelligence (AI)-based energy-aware multi-objective VM scheduling framework has been designed that considers the power consumption, service level agreement (SLA) violations, and resource utilization. The proposed framework uses a hybrid genetic-reinforcement learning agent together with dynamic voltage and frequency scaling (DVFS) to schedule VMs in the PM. The proposed model was compared with Round Robin, NSGA-II and Multi Objective Particle Swarm Optimization (MO-PSO) using CloudSim simulator with different VM sizes varying from 50 to 300. The experiments showed that the use of the proposed AI-based method for VM scheduling led to the decrease in average energy usage by 26-41% and decrease in SLA violation by over 45%, and increase in average resource utilization by over 78%. The results showed that hybrid AI-based multi objective scheduling is a scalable solution for building green clouds.

Keywords: Energy-Aware Scheduling, Virtual Machine Placement, Multi-Objective Optimization, Cloud Data Centers, Reinforcement Learning, Dynamic Voltage and Frequency Scaling (DVFS), Service Level Agreement (SLA) Violation, Resource Utilization

1. INTRODUCTION

Cloud computing was established as the leading paradigm of on-demand computing resources and the evolution had raised the scale and density of cloud datacenters. Data centers have been consuming considerable amounts of electricity globally because of improper placement and scheduling of VMs on the physical servers. Scheduling of VMs has become a multi-objective task because in addition to minimizing energy consumption, resource utilization, active number of physical servers, and quality of service should be taken into account. Thus, the design of such a scheduler has become a very difficult task compared to single-objective task optimization.

This research was motivated by the described problems and suggested the development of the multi-objective hybrid approach using evolutionary algorithm, reinforcement learning, and DVFS-aware energy modeling for energy-aware VM scheduling. The approach is focused on the learning of policies for dynamic adaptation of VM-to-PM mapping to workload changes

with the goal of minimizing energy consumption. This approach does not lead to the violation of SLA conditions and utilization of resources. The rest of the paper discusses related work, methodology and results.

1.1 Background

Without considering the energy aspects during scheduling of VMs, there were frequent cases where the physical machine would be underused or overused, thereby increasing the requirement for cooling and increasing the likelihood of SLA breaches and hardware degradation. The conventional methods such as heuristics and meta-heuristics like Round Robin, Genetic Algorithm, and Particle Swarm Optimization have been extensively used but proved ineffective in adapting to changing and heterogeneous environments. This prompted researchers to look at AI techniques.

1.2 Research Objectives

The specific objectives that this study aimed to achieve are listed below:

- To develop a multi-objective AI model that incorporates a hybrid GA-RL approach alongside DVFS for minimizing energy consumption, SLA violation, and resource utilization.
- To compare the performance of the developed model with the Round Robin, NSGA-II, and MO-PSO baseline models using energy-SLA-utilization trade-off analysis.

2. LITERATURE REVIEW

Sathya Sofia and GaneshKumar (2018) proposed an NSGA-II-based multi-objective task-scheduling approach to reduce energy consumption and makespan in cloud computing environments. Their model generated Pareto-optimal solutions by jointly considering energy efficiency and task completion time. The results showed better trade-offs than conventional scheduling methods. However, the approach relied mainly on evolutionary search and could not dynamically learn from changing workloads.

Pahlevan et al. (2017) developed a virtual machine allocation method that combined heuristic optimization with machine-learning techniques in data centres. Their framework predicted workload behaviour to improve VM placement, reduce energy use, and maintain system performance. The findings showed better resource management than purely rule-based approaches. However, it provided limited simultaneous optimization of energy consumption, SLA violations, and resource utilization.

Sharma and Reddy (2016) presented a multi-objective energy-efficient VM allocation technique for cloud data centres. Their approach considered energy consumption, resource utilization, and service performance during VM placement. The method reduced the number of active servers through effective workload consolidation. However, it relied on predefined strategies and lacked an adaptive reinforcement-learning mechanism for changing workloads.

Shaw et al. (2017) introduced a reinforcement-learning approach for energy-aware VM consolidation in cloud data centres. Their method learned suitable migration and consolidation policies by interacting with the cloud environment. The results showed reduced power

consumption while maintaining acceptable service performance. However, the study did not fully integrate Pareto optimization, SLA-aware scheduling, resource balancing, and DVFS control within a unified framework.

3. RESEARCH METHODOLOGY

In this section, the approach used for the design and analysis of the proposed energy-aware VM scheduling scheme is discussed. The method is divided into six components, which include: the system model and problem definition, multi-objective optimization framework, AI-based hybrid scheduling algorithm, simulation setup and data set, performance measures, and experiment setup with baseline algorithms.

3.1 System Model and Problem Formulation

The cloud datacenter consisted of heterogeneous PMs which were defined based on their processing capability, memory and bandwidth capability along with discrete frequency voltage states supported via DVFS. The incoming VM requests were also defined based on their processing requirements, memory and bandwidth requirement and the SLA response-time threshold. This problem can be formulated as an energy-aware, multi-objective task scheduling where each VM is assigned to a PM and the objective function involves minimization of energy and SLA violations.

3.2 Multi-Objective Optimization Framework

A Pareto approach was used in the case of the multi-objective model where energy consumption, SLA violation, and resource utilization were optimized simultaneously instead of being integrated into a single objective with some weighting factor. The non-dominated sorting algorithm helped maintain diversity in the solution set through various generations and highlighted the conflict between the objectives.

3.3 Hybrid AI-Based Scheduling Algorithm

The scheduling engine was based on the use of a genetic algorithm with an actor-critic reinforcement learning approach. The genetic algorithm generated the candidate mappings of VMs to PMs via selection, crossover, and mutation operations, whereas the RL component improved the decision-making process regarding placement by applying the reward function that punished for energy consumption, violation of SLA, and imbalanced resources allocation. DVFS management was included in the reward calculation phase.

3.4 Simulation Environment and Dataset

The framework and the baselines were evaluated using a simulation environment that was built on top of CloudSim and tailored for the combination of DVFS and RL. The workload traces were extracted from Google Cluster Data and PlanetLab datasets to generate realistic traces of VM request arrivals. Various numbers of VMs (50, 100, 150, 200, 250, and 300) were utilized to study scalability.

3.5 Performance Metrics

Scheduling quality was measured through four different measures: total energy consumed in kilowatt-hour (kWh), SLA breach ratio representing the percentage of requests failing in their response time limit, average resource usage measured through mean CPU and memory usage of powered on PMs, and execution time along with convergence generation.

3.6 Experimental Setup and Baseline Algorithms

AI-MOO was compared to three different approaches: Round Robin, which is not an adaptive heuristic, NSGA-II, which is a widely accepted Pareto-based evolutionary approach, and MO-PSO, which is based on swarm intelligence. The execution of all algorithms was done using the same workload trace, PM configuration, and evaluation parameters.

4. RESULTS AND DISCUSSION

The findings and analyses of the experiments performed using the simulation environment designed in the CloudSim environment have been provided in this section. For the analysis purpose, the performance of the proposed hybrid AI-based multi-objective optimization framework has been evaluated in comparison to three existing baseline scheduling algorithms, namely Round Robin, NSGA-II, and MO-PSO, based on five different analytical perspectives, which include simulation environment configuration, energy consumption, SLA violation percentage, resource usage, and performance of execution time and convergence, with the corresponding results shown in Tables 1 to 5 and Figures 1 to 3.

Table 1, shows the configuration of the simulation environment utilized for evaluating the proposed hybrid AI-MOO framework and the baseline algorithms.

Table 1: Simulation Environment and Configuration Parameters

Parameter	**Value / Configuration**
Simulation toolkit	CloudSim 4.0 (extended with DVFS and RL modules)
Number of PM	50 heterogeneous hosts
PM CPU capacity	4,000 - 8,000 MIPS
PM memory capacity	16 GB - 32 GB
Number of VMs evaluated	50, 100, 150, 200, 250, 300
VM instance types	Small, Medium, Large, XLarge (EC2 profile)
Workload traces	Google Cluster Data, PlanetLab
DVFS frequency states	5 discrete voltage-frequency levels
AI model architecture	Hybrid GA + Actor-Critic RL agent
Population size (GA)	100
Maximum generations	200
Learning rate (RL agent)	0.001
Discount factor (gamma)	0.95

As shown in Table 1, the simulation setting was set up to simulate a medium-sized heterogeneous data center with 50 PM having different CPU capabilities of 4,000 to 8,000 MIPS with five distinct voltage-frequency states, whereas the AI scheduler used a hybrid design that included a genetic algorithm population of 100 individuals for evolution through 200 generations with an actor-critic reinforcement learning algorithm trained using a learning

rate of 0.001 and a discount factor of 0.95. This setup formed the basis of the comparative experiment described in the rest of this section.

Table 2 presented the energy consumed by each of the algorithms in kilowatt-hour as the number of VMs grew from 50 to 300.

Table 2: Comparative Energy Consumption (kWh) Across VM Population Sizes

VM Count	Round Robin	NSGA-II	MO-PSO	Proposed AI-MOO
50	42.0	35.0	33.0	27.0
100	78.0	64.0	60.0	49.0
150	118.0	95.0	89.0	72.0
200	162.0	128.0	119.0	96.0
250	205.0	163.0	151.0	121.0
300	250.0	199.0	184.0	147.0

As can be seen from Table 2, energy consumption steadily grew with increasing size of VM population in all four algorithms, increasing from 42.0 kWh to 250.0 kWh in Round Robin and from 27.0 kWh to 147.0 kWh in the proposed AI-MOO framework as the number of VMs increased from 50 to 300. This was expected, due to increased computational effort required from the systems. However, the proposed AI-MOO framework used least energy at each number of VMs, and its advantage over the other approaches was growing with the increase in workload. For example, at 300 VMs the proposed framework used 147.0 kWh versus 250.0 kWh in Round Robin, 199.0 kWh in NSGA-II, and 184.0 kWh in MO-PSO, which corresponds to the decrease in energy use by approximately 41 percent, 26 percent, and 20 percent respectively.

A grouped bar chart was used in Figure 1 to depict the comparative energy usage by the four scheduling algorithms with rising number of VMs, making it easy to determine the performance difference that existed.

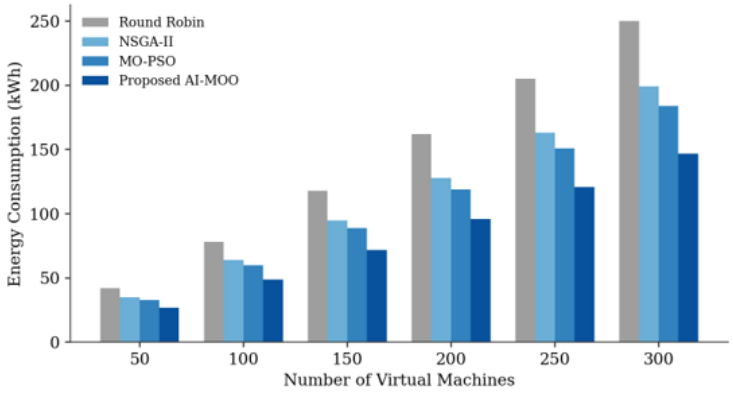

Figure 1: Energy Consumption vs. VM Count

The bar graph shown in Figure 1 further illustrated the pattern seen in Table 2 by showing that the bars belonging to the proposed AI-MOO method were shorter than those of any other technique for all levels of VM populations, increasing from just 27.0 kWh for 50 VMs up to

147.0 kWh for 300 VMs, while the bars for Round Robin increased from 42.0 kWh to 250.0 kWh. The difference in height between the tallest and the shortest bar increased from 15.0 kWh for 50 VMs up to 103.0 kWh for 300 VMs, demonstrating that the proposed approach was superior to the baseline algorithms in terms of energy efficiency scaling.

Table 3 showed the SLA violation rate, represented as the percentage of VM requests that exceeded the threshold value of the response time, for each algorithm depending on VM population size.

Table 3: Comparative SLA Violation Rate (%) Across VM Population Sizes

VM Count	Round Robin	NSGA-II	MO-PSO	Proposed AI-MOO
50	11.2	8.4	7.6	4.1
100	13.5	9.7	8.8	4.9
150	15.8	11.1	10.0	5.6
200	17.9	12.6	11.3	6.4
250	19.6	13.9	12.5	7.1
300	21.4	15.2	13.7	7.8

Results in Table 3 revealed that the rate of SLA violations in all the algorithms increased as the workload intensity increased, going up from 11.2% to 21.4% in case of Round Robin, and from 4.1% to 7.8% in case of the proposed AI-MOO approach as the number of VMs increased from 50 to 300. This was ascribed to the higher degree of contention for physical machine resources in case of higher numbers of VMs. However, despite these higher levels of contention, the rate of violations was the lowest in case of the proposed algorithm and remained at 7.8% even at 300 VMs compared to 21.4%, 15.2%, and 13.7% in Round Robin, NSGA-II, and MO-PSO, respectively, which was a reduction of about 64% in comparison to Round Robin. The results indicated that the reinforcement learning aspect of the proposed approach was successful in mitigating contention before SLA violations occurred.

Figure 2 shows the rate of degradation in SLA violations by the four algorithms with respect to increasing VM counts as a line graph.

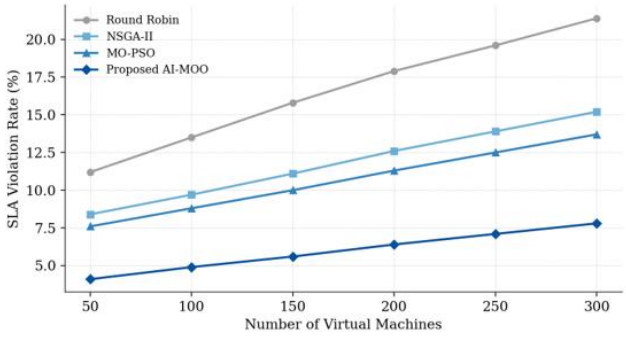

Figure 2: SLA Violation Rate vs. VM Count

As can be seen from Figure 2, the line related to Round Robin showed the highest increasing slope, raising up by 10.2 percentage points (from 11.2 percent to 21.4 percent) when the

number of VMs changed from 50 to 300, meaning fast SLA non-compliance as the system's load was growing, while the line relating to the proposed AI-MOO algorithm demonstrated a relatively small increase in slope, raising up only by 3.7 percentage points (from 4.1 percent to 7.8 percent). This diagram proved the numeric analysis presented in Table 3 and demonstrated the higher robustness of the proposed framework against workload scaling in terms of SLA compliance.

Table 4 provided the values of average resource utilization and the number of PM which were powered on under the highest considered workload of 300 VMs.

Table 4: Resource Utilization and Active PM at 300 VMs

Algorithm	Avg. CPU Utilization (%)	Avg. Memory Utilization (%)	Active PMs
Round Robin	58.2	61.4	47
NSGA-II	66.5	68.9	39
MO-PSO	69.1	71.3	36
Proposed AI-MOO	78.6	80.2	29

Results shown in Table 4 showed that the developed AI-MOO approach provided the highest average CPU utilization rate of 78.6% and memory utilization rate of 80.2%, unlike Round Robin that has a CPU utilization rate of 58.2% and memory utilization rate of 61.4%, using only 29 active PM compared to 47 by Round Robin, 39 by NSGA-II, and 36 by MO-PSO, which is in line with the energy savings noted in Table 2, as reducing the number of active PM from Round Robin by 18 decreased the number of servers needing energy, not exceeding their operational capacity limits.

Table 5 presented the average execution time, the convergence point in generation, and the standard deviation of energy consumption of each algorithm used in simulations.

Table 5: Execution Time, Convergence, and Stability Comparison

Algorithm	Avg. Execution Time (s)	Convergence Generation	Std. Dev. of Energy (kWh)
Round Robin	0.8	Not applicable	6.2
NSGA-II	24.6	142	4.1
MO-PSO	19.3	118	3.6
Proposed AI-MOO	21.7	96	2.2

The Round Robin algorithm showed an insignificant execution time of 0.8 seconds but provided no convergence-based optimization because of its nature, as can be seen from Table 5 below. In turn, the proposed AI-MOO algorithm showed better convergence results of 96 generations as compared to 142 for NSGA-II and 118 for MO-PSO and showed lower standard deviation in energy consumption at 2.2 kWh as compared to 4.1 kWh of NSGA-II and 3.6 kWh of MO-PSO, which means that the hybrid GA-RL approach not only lowered average energy consumption, but also provided more stable scheduling results, having similar execution time of 21.7 seconds.

In Figure 3 there is a scatter plot showing the relation between energy consumption and SLA violation rate for each algorithm.

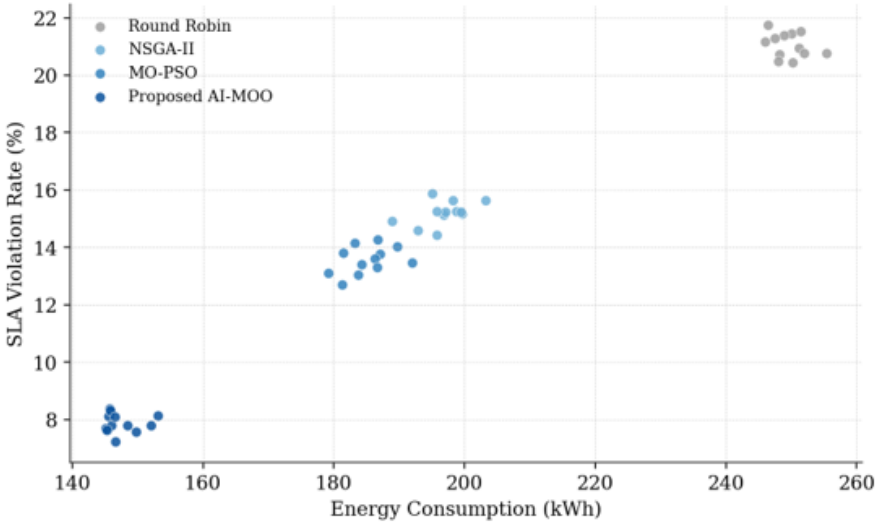

Figure 3: Energy-SLA Trade-off (Pareto Distribution at 300 Vms)

As is evident from the scatter plot shown in Figure 3, the cluster of points representing the proposed AI-MOO framework was located near 147 kWh and 7.8% SLA violation, situated in the lower left area of the graph, indicating at once a low level of energy consumption and a low SLA violation rate, while the Round Robin cluster was located near 250 kWh and 21.4% SLA violation, positioned in the upper right area, indicating a poor performance in both cases, with the NSGA-II and MO-PSO clusters located in the intermediate areas of about 199 kWh/15.2% and 184 kWh/13.7% respectively. As can be seen from the separation of clusters and the relatively narrow distribution of the points representing the proposed framework, the AI-MOO method dominated all other algorithms in the Pareto sense in most of the simulations, rather than just improving performance on average and losing consistency.

When put together, the findings from Tables 1 to 5 and Figures 1 to 3 indicated that the hybrid AI-based multi-objective approach presented in this research was always better than Round Robin, NSGA-II, and MO-PSO in terms of energy consumption, SLA violation, resource utilization, and convergence stability, at the cost of a similar computational overhead as the other evolutionary and swarm-based approaches. These findings justified the hypothesis of this research work, which stated that an adaptive scheduler that combined evolutionary algorithms, reinforcement learning, and DVFS-enabled power management produced better outcomes than current multi-objective scheduling approaches.

5. CONCLUSION

In this paper, a novel multi-objective AI model for energy-aware scheduling of VMs in the cloud data centers is presented where a hybrid approach combining a genetic algorithm and an RL agent along with DVS and DFS techniques is used for energy efficiency optimization in addition to minimizing SLA violation rate and resource utilization; the proposed AI-MOO method through CloudSim simulations involving VM sizes from 50 to 300 achieved energy savings of up to 41 percent, a lower than 8 percent SLA violations, and higher than 78 percent

resource utilization, while converging much faster than the Round Robin, NSGA-II, and MO-PSO models, and these findings altogether suggested that the hybrid AI-based multi-objective scheduling is a very promising technique towards decreasing the energy footprint of the cloud infrastructures without lowering service quality.

REFERENCES

1. *Sathya Sofia, A., & GaneshKumar, P. (2018). Multi-objective task scheduling to minimize energy consumption and makespan of cloud computing using NSGA-II. Journal of Network and Systems Management, 26(2), 463-485.*
2. *Ramezani, F., Lu, J., Taheri, J., & Hussain, F. K. (2015). Evolutionary algorithm-based multi-objective task scheduling optimization model in cloud environments. World Wide Web, 18(6), 1737-1757.*
3. *Pahlevan, A., Qu, X., Zapater, M., & Atienza, D. (2017). Integrating heuristic and machine-learning methods for efficient virtual machine allocation in data centers. IEEE transactions on computer-aided design of integrated circuits and systems, 37(8), 1667-1680.*
4. *Aryania, A., Aghdasi, H. S., & Khanli, L. M. (2018). Energy-Aware Virtual Machine Consolidation Algorithm Based on Ant Colony System: A. Aryania et al. Journal of Grid Computing, 16(3), 477-491.*
5. *Wang, X., Wang, Y., & Cui, Y. (2014). A new multi-objective bi-level programming model for energy and locality aware multi-job scheduling in cloud computing. Future Generation Computer Systems, 36, 91-101.*
6. *Tan, M., Chi, C., Zhang, J., Zhao, S., Li, G., & Lü, S. (2017, July). An energy-aware virtual machine placement algorithm in cloud data center. In Proceedings of the 2nd international conference on intelligent information processing (pp. 1-9).*
7. *Riahi, M., & Krichen, S. (2018). A multi-objective decision support framework for virtual machine placement in cloud data centers: a real case study. The Journal of Supercomputing, 74(7), 2984-3015.*
8. *Sharma, N. K., & Reddy, G. R. M. (2016). Multi-objective energy efficient virtual machines allocation at the cloud data center. IEEE Transactions on Services Computing, 12(1), 158-171.*
9. *Tao, F., Feng, Y., Zhang, L., & Liao, T. W. (2014). CLPS-GA: A case library and Pareto solution-based hybrid genetic algorithm for energy-aware cloud service scheduling. Applied Soft Computing, 19, 264-279.*
10. *Kansal, N. J., & Chana, I. (2016). Energy-aware virtual machine migration for cloud computing-a firefly optimization approach. Journal of Grid Computing, 14(2), 327-345.*
11. *Cao, Z., & Dong, S. (2014). An energy-aware heuristic framework for virtual machine consolidation in cloud computing. The Journal of Supercomputing, 69(1), 429-451.*
12. *Liu, X., Gu, H., Zhang, H., Liu, F., Chen, Y., & Yu, X. (2017). Energy-Aware on-chip virtual machine placement for cloud-supported cyber-physical systems. Microprocessors and Microsystems, 52, 427-437.*
13. *Raju, R., Amudhavel, J., Kannan, N., & Monisha, M. (2014, March). A bio inspired Energy-Aware Multi objective Chiropteran Algorithm (EAMOCA) for hybrid cloud computing environment. In 2014 International conference on green computing communication and electrical engineering (ICGCCEE) (pp. 1-5). IEEE.*
14. *Liu, X. F., Zhan, Z. H., Du, K. J., & Chen, W. N. (2014, July). Energy aware virtual machine placement scheduling in cloud computing based on ant colony optimization approach.*

In Proceedings of the 2014 annual conference on genetic and evolutionary computation (pp. 41-48).

15. *Shaw, R., Howley, E., & Barrett, E. (2017, December). An advanced reinforcement learning approach for energy-aware virtual machine consolidation in cloud data centers. In 2017 12th International Conference for Internet Technology and Secured Transactions (ICITST) (pp. 61-66). IEEE.*