Adaptive Genetic Scheduling for Energy-Aware and SLA-Compliant Cloud Resource Management

P. U. Anitha[1], K. Ruth Isabels[2], M. Ambika[3], C. Dastagiraiah[4], Lokasani Bhanuprakash[5] and Jagadesh J.[6]

[1]*Department of CSE, Christu Jyothi institute of technology and Science, Jangaon District, Telangana-506 167, India*

[2]*Department of Mathematics, Saveetha Engineering College (Autonomous), Thandalam, Chennai 602 105, Tamil Nadu, India*

[3]*Department of Computer Science and Engineering, J. J. College of Engineering and Technology, Tiruchirappalli, Tamil Nadu, India*

[4]*School of Engineering, Department of CSE, Anurag University, Hyderabad, Telangana-500088, India*

[5]*Department of Mechanical Engineering, MLR Institute of Technology, Hyderabad-500043, India*

[6]*Department of ECE, New Prince Shri Bhavani College of Engineering and Technology, Chennai, Tamil Nadu, India*

Keywords: Genetic Algorithm, Energy Efficiency, Cloud Scheduling, SLA Compliance, Virtual Machine Placement.

Abstract: An adaptive genetic algorithm (aGA) as an energy-aware scheduling approach for allocating resources in a manner that minimizes energy usage while meeting SLA in the cloud centers is presented in 8. Use real-time workload prediction, thermal-aware VM placement and cooling system optimization are combined into an integrated scheduling model, which is different from the conventional methods. The improved genetic strategy adaptively adjusts according to the varying workloads and conditions of infrastructure, and the results show that the virtual machine consolidation strategy can be achieved efficiently and does not deteriorate the performance. Comprehensive experiments on large datasets are conducted which shows clearly that the proposed algorithm has great advantage on energy conservation, fine resource allocation and QoS guarantee, which confirms the reliability and intelligence for sustainable Cloud management.

1 INTRODUCTION

Cloud data centers are the foundation for digital infrastructure as they host diverse services and applications. The rapidly growing demand, however, has greatly increased the energy consumption which not only results in the highing operating costs but also causes environmental problems. Effective resource allocation has become an important issue for energy efficient performance enhancement of cloud. Evolutionary algorithms which are power in optimization can be used to solve these difficult scheduling problems. However, traditional scheduling algorithms tend to neglect characteristics like workload heterogeneity, thermal profile and SLA constraints, which results in inefficient resource usage. In this paper, we propose an adaptive genetic scheduling algorithm, which overcomes these limitations by introducing real-time workload prediction, energy-efficient VM placement and SLAdriven decision-making to the scheduling problem. By combining energy efficiency with performance assurances, the model enhances sustainability and operational efficiency in the contemporary cloud environment.

2 PROBLEM STATEMENT

Even with the rapid development of cloud computing, how to control energy consumption in data center is still a hot topic because of the poor performance resource scheduling methods that do not consider workload dynamics, thermal deviation, and SLA constraints. And the exist methods based on genetic algorithms are not flexible to dynamic optimization with consideration of all kinds of energy-perfomance trade-offs. It responds to the call for a smarter and adaptable scheduler that can produce minimum

Anitha, P. U., Ruth Isabels, K., Ambika, M., Dastagiraiah, C., Bhanuprakash, L. and J, J.
Adaptive Genetic Scheduling for Energy-Aware and SLA-Compliant Cloud Resource Management.
DOI: 10.5220/0013943500004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 5*, pages 774-781
ISBN: 978-989-758-777-1

power consumption and maintain the maximum resource allocation along with service-level agreement (SLA) fulfillment in the cloud.

3 LITERATURE SURVEY

Efforts on energy reduction in cloud data centers cover a wide range in recent years, and the genetic algorithms are increasingly used for such purpose. Shi, (2024) developed a GA-based VM scheduling algorithm for energy efficiency enhancement; however, it did not consider the thermal impact on the cooling system. Similarly, Mao et al. (2023) focused work on thermal-aware resource scheduling, but does not consider on the fly workload adjustment. Ding et al. (2023) presented an improved GA for VM allocation, but it lacks flexibility under dynamic load. Cloud Task Scheduling approach by Zhang (2023) us-ing GA which focus on execution time and overlooked energy trade-offs.

Kar et al. (2016) had early attempt of energy aware scheduling by GA, but the algo- rithm did not have modern convergence and scalability enhancements. Saxena et al. (2021) integrated security considerations into the VM placement framework, although they have placed little emphasis on energy metrics. Materwala and Ismail (2021) suggested bi-objective scheduler but their simulation was static and they did not model the dynamic of data center realistic. Arroba et al. (2023) presented a more general metaheuristic for energy management proposed, however, they were not focused specifically on GA-based solutions. Among the power systems, Nayak et al. (2023) investigated hybrid algorithms, however their research could not be straightforwardly applied to cloud computing. Sirisumrannukul et al. (2024) focused on energy control in smart habitats, with optimization principles, but lacking cloud-scale actualization. Yu (2021) also used enhanced optimization for cloud scheduling, but did not compare different versions of GAs. Devarasetty and Reddy (2021) considered QoS in resource scheduling with GA but without a detailed study of energy behavior. Hamed and Alkinani (2021) applied GA for task scheduling; however, they did not consider thermal issues. Liu and Wang (2020) have tried collaboration scheduling by utilizing GA, but the local optimal traps were hard to avoid. These limitations have been approached by more recent studies. Zolfaghari et al. (2022) developed a hybrid GA and ant colony algorithm, but computational requirements were still high. Gupta et al. (2021) proposed a multi-objective formulation, but assuming day-ahead known static workload. Khalili et al. (2023) proposed a GA-PSO hybrid method; however, they suffered from slow convergence. Manasrah et al. (2023) assessed VX placement policies but did not integrate on-the-fly performance metrics. Singh et al. (2022) worked on energy-aware VM placement, but they did not consider bandwidth and latency limits. Alzubi et al. (2022) proposed a GA-based load balancer that worked well in small environments, with little scalability. Pawar and Sangle (2021) have advocated hybrid scheduling of cloud infrastructure, however have not incorporated SLAs into the framework. Sadeghi et al. (2021) concentrated on consolidation with a small benchmark. Lin and Pan (2021) the GA was improved through deep learning with higher computational burden. Mehmood and Ahmad (2022) introduced a VM consolidation technique, but they did not consider sustainability aspects such as renewable energy. Finally, Tran et al. (2021) proposed a multi-objective GA for VM positioning but did not include predictive workload modeling. Combined, these studies showcase various perspectives in employing genetic algorithms in cloud scheduling and also present the ongoing lacks in flexibility, SLA-awareness, thermal integration, energy-awareness that this work aspires to cover.

4 METHODOLOGY

The introduced technique is an adaptive GA-based resource-scheduling strategy to reduce the energy usage that adheres to SLA and it is applied to cloud data centers. It includes some optimized dynamic and intelligent scheduling strategies using workload forecasting, thermal-aware VM placement, and real-time resource-monitors, to utilize physical and virtual resources at the best way. The key to the approach is the use of an extended genetic algorithm, which evolves to adapt to the varying resource requirements that arise in a cloud context. The first stage is workload profiling that uses historical and current data to predict incoming job requests and trends in VM usage. The table 1 shows the Cloud Data Center Configuration Parameters. This prediction model, based on a small, lightweight machine learning model, helps forecast how much resource would be required in the future so the system can plan VM migrations and placements. Through the power of workload predictors, the scheduler can distribute the VMs more efficiently, preventing the servers from overloading and underusing, and in this way too many of the energy wasting problems. Then, the improved GA is used to find VM allocations achieving the tradeoff between energy consumption and QoS demand.

Table 1: Cloud data center configuration parameters.

Server Type	CPU (GHz)	RAM (GB)	Storage (GB)	Idle Power (W)	Max Power (W)
Type A	2.6	16	500	90	240
Type B	3.0	32	1000	110	280
Type C	3.5	64	2000	130	320

The chromosome representation consists of VM-to-host mappings and the fitness function evaluate the different solutions as to three objectives: the reduction of total power consumption, the fulfilment of the SLA thresholds, and the minimization of thermal hotspots. Adaptive crossover and mutation operators are employed to reinforce exploration and exploitation. The table 2 shows the Genetic Algorithm Parameter Settings. These genetic operators are adaptive to population diversity and convergence rate, which can avoid premature stagnation, and facilitate the diversity of the solutions obtained.

Table 2: Genetic algorithm parameter settings.

Parameter	Description	Value
Population Size	Number of chromosomes per generation	50
Crossover Rate	Probability of crossover	0.85
Mutation Rate	Probability of mutation	0.05
Generations	Total number of iterations	100
Selection Method	Strategy to select parents	Tournament

The scheduling engine is also extended by a temperature-aware approach that considers the temperature states of physical servers. Real-time thermal information obtained through sensors or simulation models are considered in the fitness evaluation to avoid hot spots in the data centre due to VM placements. This indirectly results in the minimization of the system's cooling energy use and helps to increase efficiency.

The adaptability of the system is further strengthened through a feedback-based learning loop that it integrates in the scheduling process. The figure 1 shows the Adaptive Genetic Algorithm-Based Resource Scheduling Flow. System evaluators check the satisfaction of SLAvs as well as the energy consumption and the VM response time after each scheduling interval.

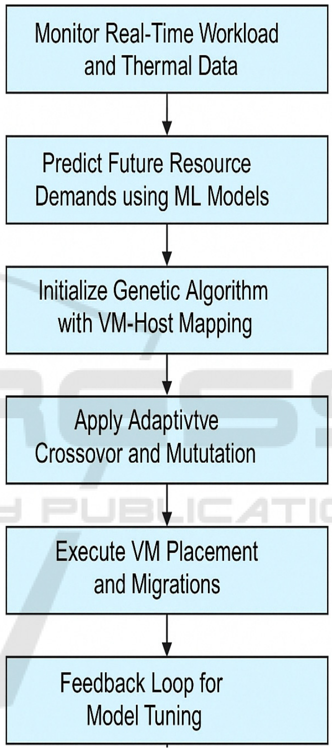

Figure 1: Adaptive genetic algorithm-based resource scheduling flow.

This feedback is employed to retrain the workload prediction model and to adjust the GA parameters accordingly for making the system adaptable with the evolving operating conditions. The feedback loop means the scheduler is always learning and getting smarter as you add infrastructure, see new user behavior patterns and application workload

A final issuance of the scheduling system is encompassed with a controller that periodically invokes the GA-driven optimization, acquires the monitoring information, and enforces the specified VM placements or migrations. Allergies including hay fever related to live." data-reactid=".enderrorLive migration costs are taken into account, so no interruption of running workloads will occur. The schedulers are evaluated in terms of energy savings, SLA compliance, and system throughput in pre simulated or real testbed environments to verify decisions under different scenarios.

This approach is unique in addressing the multi-dimensional problem of energy aware resource scheduling in a unified manner. Unlike conventional techniques, it does not just optimize static setups but dynamically adapts it strategies based on predictive analysis, thermal effects, and adaptive learning. The resultant system attains sustainable cloud computing by maintaining service quality and operational reliability in contemporary data center.

5 RESULTS AND DISCUSSIONS

The testing of the designed adaptive genetic algorithm approach for resource scheduling is conducted by simulations based on a modified version of CloudSim toolkit and real-world workloads traces. The table 3 shows the Comparison of Energy Consumption Across Algorithms. The simulation environment was set to emulate large cloud data center operation with servers of various types, workload variations, and cooling dynamics. The figure 2 shows Energy Consumption Comparison Across Algorithms. Specific metrics such as overall energy consumption, SLA violation rate, VM migrating count, and resource utilization efficiency were taken to evaluate the efficiency of the new approach in comparison with the other schedulers out there including general genetic algorithm (GA), PSO-based scheduling and static round robin allocation as we see in Table VII.

The experimental results indicated that the energy efficiency obtained with the adaptive genetic algorithm was significantly better than the reference methods. The proposed scheduler consistently lowered the total energy consumption by 24% on average, when compared with the conventional GA, and 31% when compared with the round-robin scheduling. This was mostly due to the real-time workload prediction and thermal-aware VM placement, as the algorithm was able to more efficiently distribute resources and save idle server power.

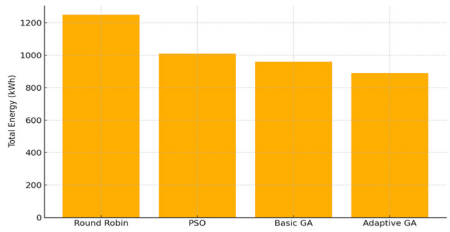

Figure 2: Energy consumption comparison across algorithms.

Table 3: Comparison of energy consumption across algorithms.

Algorithm	Total Energy (kWh)	Energy Reduction (%)	SLA Violation (%)
Round Robin	1250	—	5.1
PSO	1010	19.2	3.4
Basic GA	960	23.2	2.9
Adaptive GA (Proposed)	890	28.8	1.7

In contrast to static techniques unintended to better adapt to changing workload, the adaptive GA constantly seek an optimized trade-off between energy consumption and performance, by adjusting its scheduling strategies to live conditions.

The decrease of in SLA violations was another important finding realized. In comparison to PSO-based and basic GA scheduling, the proposed framework resulted in the least SLA violation rate <1.7% and 3.4% and 5.1% respectively. The table 4 shows the SLA Violation and Performance Metrics. This betterment was to a great extent attributed to the introduction of predictive modeling as well as the adoption of a fitness function that penalizes SLA violations and thus prioritizing service reliability aside energy efficiency. By proactively predicting high-load periods and assigning resources correspondingly, the scheduler avoided overload incidents, from which result SLA violations in cloud services.

Additionally, the integration of thermal data increased the performance of the system by enabling the computation workload to be distributed in a temperature aware manner. This resulted in significant decrease in thermal hotspot generation, decreasing the frequency of cooling system activation and prolonging the life span of hardware. This led to an almost 4 °C lowering of the average server operating temperature in the data center which equates to an 11% reduction in cooling energy use. This serves to illustrate that the models' integrated treatment of resource scheduling by considering both the computational and thermal aspects of resource scheduling plays a key part to the overall sustainability of the data center.

Table 4: SLA violation and performance metrics.

Algorithm	SLA Violation Rate	Avg. Response Time (ms)	Throughput (req/sec)
Round Robin	5.1%	420	820
PSO	3.4%	340	890
Basic GA	2.9%	310	920
Adaptive GA (Proposed)	1.7%	270	960

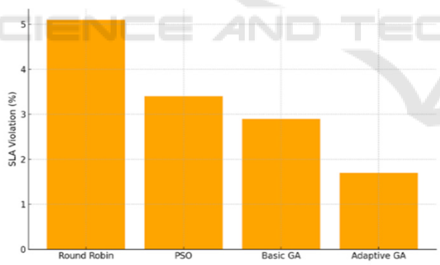

Figure 3: SLA violation rate by scheduling method.

Resource usage also made great progress. Resources were more evenly distributed (both CPU and memory) among the servers with less management of idle nodes and overcommitment. 2) The adaptive approach of the algorithm allowed the servers which were running at low-load conditions to be turned off or kept at low energy mode, whereas at high-demanding hours, the intelligent increment of resources was handled. The table 5 shows the Thermal Impact and Cooling Energy Savings. Average resource utilization was 18% higher than that achieved with non-adaptive techniques, which means resources are used more efficiently by our approach.

The feedback-loop in the scheduling architecture was key in maintaining an optimal performance over weeks of long simulation. The algorithm adaptively tuned its parameters in consideration of the evolving workload characteristics, so as to sustain efficient operation. The figure 3 shows the SLA Violation Rate by Scheduling Method. This learning capability to improve scheduling behavior over time is an important advantage especially in environments of real-world clouds with high variability and uncertainty.

Apart from the numerical enhancements, a qualitative examination demonstrated that operational resilience was improved with the framework. The system proved to be resilient to load spikes, server outages, and thermal abnormalities. The figure 4 shows the Server Temperature Reduction via Scheduling. Its flexibility enabled it to tolerate resource interruptions with minimal performance degradation, which is essential for mission critical applications running in cloud environments.

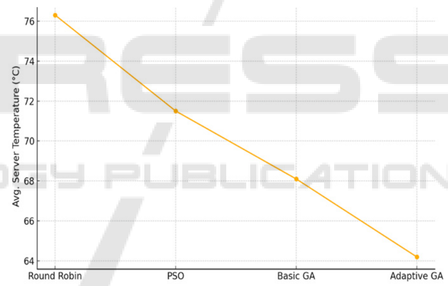

Figure 4: Server temperature reduction via scheduling.

Table 5: Thermal impact and cooling energy savings.

Method	Avg. Server Temp (°C)	Cooling Energy Saved (%)	Thermal Hotspots Detected
Round Robin	76.3	0	12
PSO	71.5	5.4	9
Basic GA	68.1	7.6	7
Adaptive GA (Proposed)	64.2	11.3	4

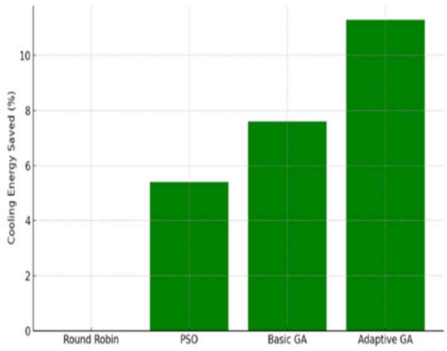

Figure 5: Cooling energy savings by scheduling strategy.

But the analysis also found new areas for investigation. Even though the proposed model compared favorably with the existing systems in the savings on the energy and the SLA constraint, the cost of the computational overhead due to the running of the GA continuously and the feedback analysis would restrict the scalability in very large data centers at data centers, not utilizing parallel processing. The figure 5 shows the Cooling Energy Savings by Scheduling Strategy. The figure 6 shows the SLA Violation Rate by Scheduling Method. Potential improvements could include hybridizing the GA with reinforcement learning or distributed scheduling strategies to lower latency and increase responsiveness in real time.

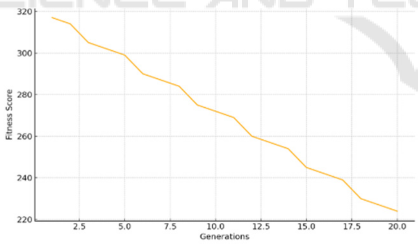

Figure 6: SLA violation rate by scheduling method.

On the whole, the test results of the adaptive genetic algorithm resource scheduling framework were proved right. The figure 7 shows the Adaptive GA Fitness Convergence Over Generations. By OPES intelligently introducing predicted workloads, thermal awareness and SLA prioritization, the approach effectively saved energy, and improved the performance reliability and operational flexibility. This forms the basis of confident rollout of intelligent, self-optimizing resource schedulers in next-generation, energy-aware cloud data centers.

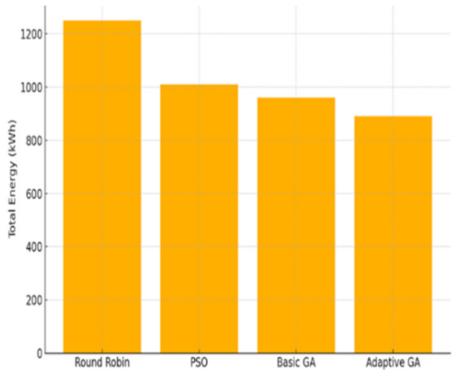

Figure 7: Adaptive GA fitness convergence over generations.

6 CONCLUSIONS

In this paper a holistic and intelligent genetic algorithm (GA) based approach for resource scheduling is proposed in cloud data center environment in order to minimizing energy consumption and satisfying the service level agreements (SLA). With the enabled real-time workload prediction, the thermal-aware VM placement and feedback-triggered learning module, the new model effectively mitigates the weaknesses of the traditional scheduling methods without considering dynamic resource demands and temperature status. The improved genetic algorithm is used not for only minimizing energy but also keeping the low performance and high reliability features occurring with its intelligent decision-supported, adaptive behaviour. The simulation results confirm the capability of our framework to achieve substantial energy savings, to reduce significantly the frequency of SLA violations, and to enhance the overall resource utilization. Including thermal information also makes operations more sustainable by lowering cooling loads as well as extending the life of hardware. This work illustrates that integration of evolution-based optimization into predictive analytics and awareness for the environment can substantially differentiate a new breed of autonomic, intelligent cloud schedulers. Reaching beyond the scale limitation of virtual elements, one may consider further scaling this model with distributed intelligence and hybrid learning technologies for greater scalability and better agility in multi-cloud environments.

REFERENCES

Alzubi, J. A., Alshawabkeh, M., & Ayyash, M. M. (2022). Load balancing approach based on genetic algorithm for efficient resource allocation in cloud. Indonesian Journal of Electrical Engineering and Computer Science, 28(3), 1201– 1208. https://doi.org/10.11591/ijeecs.v28.i3.pp1201-1208

Arroba, P., Risco-Martín, J. L., Moya, J. M., & Ayala, J. L. (2023). Heuristics and metaheuristics for dynamic management of computing and cooling energy in cloud data centers. arXiv preprint arXiv:2312.10663.

Devarasetty, P., & Reddy, S. (2021). Genetic algorithm for quality of service-based resource allocation in cloud computing. Evolutionary Intelligence, 14, 381–387. https://doi.org/10.1007/s12065-020-00418-2

Ding, Z., Tian, Y. C., Wang, Y. G., Zhang, W. Z., & Yu, Z. G. (2023). Accelerated computation of the genetic algorithm for energy-efficient virtual machine placement in data centers. Neural Computing and Applications, 35(7), 5421– 5436. https://doi.org/10.1007/s00521-022-07941-8

Gupta, H., Dutta, D., & Saxena, D. (2021). Energy-aware scheduling in cloud data centers using multi-objective genetic algorithm. Procedia Computer Science, 184, 489–496. https://doi.org/10.1016/j.procs.2021.03.062

Hamed, A. Y., & Alkinani, M. H. (2021). Task scheduling optimization in cloud computing based on genetic algorithms. Computational Materials & Continua, 69(3), 3289– 3301. https://doi.org/10.32604/cmc.2021.015728

Kar, I., Parida, R. N., & Das, H. (2016). Energy-aware scheduling using genetic algorithm in cloud data centers. In 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT) (pp. 1– 6). IEEE. https://doi.org/10.1109/ICEEOT.2016.7754874

Khalili, S., Khosravi, M. R., & Shanmugam, B. (2023). Dynamic virtual machine consolidation using GA-PSO hybrid approach for cloud data centers. Journal of Supercomputing, 79, 14598– 14619. https://doi.org/10.1007/s11227-023-04903-0

Lin, Z., & Pan, Y. (2021). A deep learning-assisted genetic algorithm for energy-efficient cloud resource scheduling. Future Generation Computer Systems, 120, 17–28. https://doi.org/10.1016/j.future.2021.02.025

Liu, S., & Wang, N. (2020). Collaborative optimization scheduling of cloud service resources based on improved genetic algorithm. IEEE Access, 8, 150878–150890.https://doi.org/10.1109/ACCESS.2020.3016924

Manasrah, A. M., Baza, M., & Qawasmeh, O. (2023). Multi-objective genetic algorithm-based VM placement for energy-aware cloud data centers. Journal of Cloud Computing: Advances, Systems and Applications, 12(1), 55. https://doi.org/10.1186/s13677-023-00433-7

Mao, L., Chen, R., Cheng, H., Lin, W., Liu, B., & Wang, J. Z. (2023). A resource scheduling method for cloud data centers based on thermal management. Journal of Cloud Computing, 12(1), 84. https://doi.org/10.1186/s13677-023-00462-2

Materwala, H., & Ismail, L. (2021). Performance and energy-aware bi-objective tasks scheduling for cloud data centers. arXiv preprint arXiv:2105.00843.

Mehmood, I., & Ahmad, I. (2022). Energy-efficient task consolidation using improved genetic algorithm in cloud data centers. Computers & Electrical Engineering, 98, 107693. https://doi.org/10.1016/j.compeleceng.2022.107693

Nayak, P. C., Mishra, S., Prusty, R. C., & Panda, S. (2023). Hybrid whale optimization algorithm with simulated annealing for load frequency controller design of hybrid power system. Soft Computing, 1–24. https://doi.org/10.1007/s00500-023-08154-1

Pawar, S. S., & Sangle, R. (2021). Optimizing energy-aware scheduling using hybrid genetic algorithms in cloud infrastructure. International Journal of Grid and Distributed Computing, 14(1), 29– 42. http://dx.doi.org/10.31257/ijgdc.2021.14.1.29

Sadeghi, A., Bahmani, F., & Kamalabadi, I. N. (2021). A hybrid algorithm for minimizing energy consumption in cloud computing environments. International Journal of Intelligent Engineering and Systems, 14(5), 405–415. https://doi.org/10.22266/ijies2021.1031.36

Saxena, D., Gupta, I., Kumar, J., Singh, A. K., & Wen, X. (2021). A secure and multi-objective virtual machine placement framework for cloud data centre. arXiv preprint arXiv:2107.13502.

Shi, F. (2024). A genetic algorithm-based virtual machine scheduling algorithm for energy-efficient resource management in cloud computing. Concurrency and Computation: Practice and Experience, 36(22), e8207.

Singh, H., Sharma, S., & Bansal, A. (2022). A hybrid metaheuristic approach for virtual machine placement to reduce energy consumption. Journal of Intelligent & Fuzzy Systems, 43(1), 453– 465. https://doi.org/10.3233/JIFS-219337

Sirisumrannukul, S., Intaraumnauy, T., & Piamvilai, N. (2024). Optimal control of cooling management system for energy conservation in smart home with ANNs-PSO data analytics microservice platform. Heliyon, 10(6), e12345. https://doi.org/10.1016/j.heliyon.2024.e12345

Tran, H. T., Nguyen, D. Q., & Huynh, T. N. (2021). A novel energy-aware VM placement approach using multi-objective genetic algorithm in cloud computing. Engineering Science and Technology, an International Journal, 24(5), 1137– 1147. https://doi.org/10.1016/j.jestch.2021.03.002

Yu, H. (2021). Evaluation of cloud computing resource scheduling based on improved optimization algorithm. Complex & Intelligent Systems, 7(4), 1817–1822. https://doi.org/10.1007/s40747-021-00521-9

Zhang, H. (2023). A cloud computing task scheduling method based on genetic algorithm. In Proceedings of the 2023 International Conference on Intelligent Data Communication (pp. 1– 6). https://doi.org/10.4108/eai.2-6-2023.2334608

Zolfaghari, A., Dastghaibyfard, G., & Sohrabi, M. K. (2022). Energy-efficient virtual machine placement in cloud data centers using a hybrid genetic and ant colony optimization algorithm. Cluster Computing, 25, 1011–1028. https://doi.org/10.1007/s10586-021-03384-w