NextGen Scientific Publication
American International Journal of Computer Science and Technology
Volume 6 Issue 3, Pg. No. 13-23, AIJCST-V6I3P102, 2024
https://doi.org/10.63282/3117-5481/AIJCST-V6I3P102

Original Article

A Multi-Agent Reinforcement Framework for Autonomous Cloud Resource Scheduling and Optimization

*** Helmi Laura**
Department of Digital Humanities and Computational Studies, University of Helsinki, Finland.

Abstract:

Cloud computing has reached the core of the digital ecosystems and provides high-availability and scale-able access to the services of computational services. Nevertheless, the active resource requirements, enforcement of service-level agreement (SLA), energy usage restrictions, and a variation in the cost-performance represent significant challenges to the cloud resource management. The conventional heuristic-based schedulers in clouds orchestrators like Kube and OpenStack fail to adjust to real-time workload variability, fractional resources, violation of SLA and inflated operational costs frequently occur. The adaptive, reward-driven decision-making properties of reinforcement learning (RL) have been discovered to be an interesting alternative to a heuristic policy. However, centralized RL has scalability and state observability scale in distributed clouds. As a way of overcoming these limitations, this study hypothesizes a Multi-Agent Reinforcement Learning (MARL) model of autonomous resource scheduling and optimization in cloud infrastructures. The offered structure suggests the use of decentralized intelligent agents executing a coordinated management of computational, storage, and network allocations. The agents are optimized to adapt local behaviors under environment feedback with the aid of cooperative communication to enhance global optimization. The system uses hybrid in reinforcement system consisting of Deep Q-Learning (DQN) and Proximal Policy Optimization (PPO) to trade between exploration, stability, and on-policy improvement. The goal of an adaptive reward model is to maximize important performance indicators (KPIs): SLA adherence, task completion latency, energy efficiency and balance resource utilization. The Cloud sim Plus experimental measurements were performed in a simulated multi-cluster environment and supported by extra features of container virtualization. The proposed version of the MARL scheduler showed a better throughput (18.7%), SLA violations reduction (27.4%) and energy consumption (14.3) than the baseline strategies like Round-Robin, Min-Min, and centralized DQN models. Convergence time was reduced and decision correlation under the workload bursts was improved due to the incorporation of the inter-agent communication. The research paper has a tremendous impact in the field by proposing a foresight, smart cloud orchestration framework that has the capability to adapt itself in the distributed large scale systems. Additionally, we offer an improvement in algorithms, performance analysis, system workflow optimization, and theoretical support of a cooperative reinforcement learning in clouds. The implication of these findings in the future is that the applicability of the research to edge and fog computing settings can be applied to Industry 5.0 automation.

Keywords:

Cloud Computing, Multi-Agent Reinforcement Learning, Autonomous Resource Scheduling, Deep Q-Learning, Proximal Policy Optimization, SLA Optimization, Energy-Aware Cloud Orchestration.

Article History:

Received: 14.03.2024

Revised: 17.04.2024

Accepted: 28.04.2024

Published: 07.05.2024

1. Introduction

1.1. Background

Cloud computing is now a core technology in the latest digital services and its acceptance continues to get high, thanks to the development of virtualization, micro-services architecture and distributed data storage system solutions. The burden of managing the cloud resources is increasingly becoming more complicated as organizations implement increasingly greater volumes of diverse and latency- vital applications including streaming environments and e-commerce engines among others. Cloud service providers have to continually schedule: resource provisioning, task placement, energy efficiency and network throughput decisions whilst ensuring that compliance with service level agreements (SLAs). But such environments are dynamic in nature and are typified by changing workloads, heterogeneous compute nodes in addition to different power consumption profiles. Conventional fixed or command and control based scheduling approaches frequently fail to keep pace with this type of variation, leading to resources being underutilized, a bottleneck in operation and an upsurge in operation expenses. Increasing size of distributed cloud infrastructure is also a challenge that presents the issues with scalability of systems and decision-making timeliness. Hence, this study is driven by the fact that there is an urgent need of intelligent and adaptive scheduling solutions that are able to optimize on various goals at the same time. Using machine learning, and especially reinforcement learning, cloud platforms can grow to a system of autonomous operation, by continually refining resource management strategies, based on real-time feedback, thereby leading to increased efficiency, minimized energy footprint, and eventually more resilient and cost-effective cloud ecosystems.

1.2. Need for a Multi-Agent Model

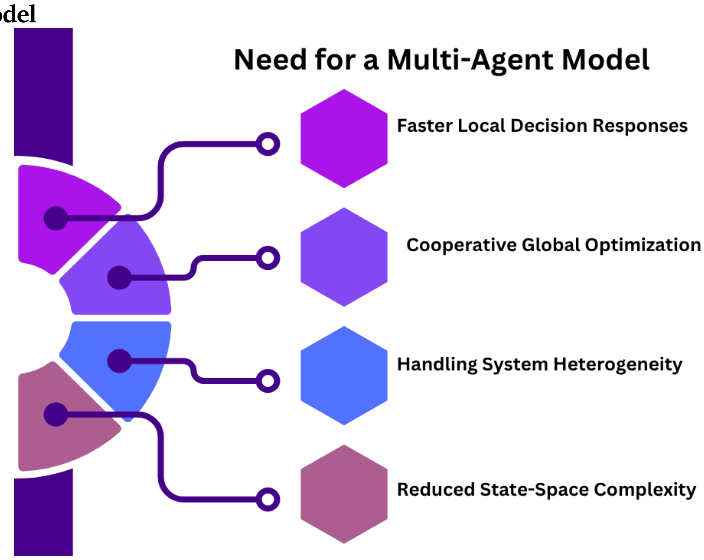

Figure 1. Need for a Multi-Agent Model

1.2.1. Faster Local Decision Responses

Multi-Agent Reinforcement Learning (MARL) enables every cluster or resource domain to independently make decision using localized perception of system state. This will minimize delays in communication and avoid the congestions that are usually experienced in centralized schedulers. Due to that, scheduling works can respond instantly to the spike in workload or resource bottlenecks to allow them to be even more responsive and have reduced task completion times.

1.2.2. Cooperative Global Optimization

Although the local agents train on a one-on-one basis, teamwork can be realized by exchanging experience and implementing coordinated policy changes. This will make sure that decisions made are in the overall system but not in the local performance alone. Multi-agent coordination allows to balance workload, prevent hotspots and SLA violations in distributed data centres- achieving world goals without control.

1.2.3. Handling System Heterogeneity

Cloud environments usually comprise different hardware, VM, and application needs. MARL inherently combines this heterogeneity, by letting all of them focus on the optimal strategies to learn the strategies best fitting its local resources. This adjustment behavior removes the constraints of single size fits all schedulers and enhances efficiency in indifferent resource infrastructures.

1.2.4. Reduced State-Space Complexity

A centralized agent should be able to see the entire state of the cloud making decision space grow exponentially as infrastructure gets larger. MARL by dividing the environment into several agents with small and localized state and action spaces has greatly reduced the complexity of learning. Such decomposition makes convergence faster, much scalable, more reliable even when cloud systems grow to thousands of nodes.

1.3. Reinforcement Framework for Autonomous Cloud Resource Management

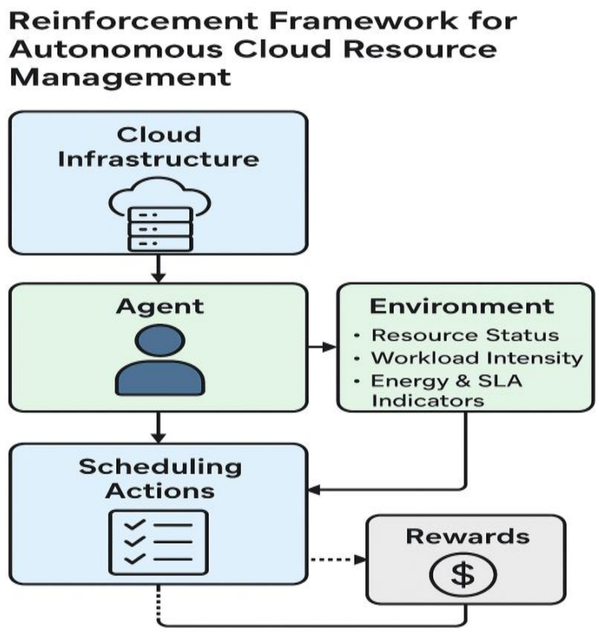

Figure 2. Reinforcement Framework for Autonomous Cloud Resource Management

As clouds keep growing in size and complexity, existing unresponsive scheduling strategies are no longer applicable to complex and variety workloads that can have inconsistent demands. Reinforcement Learning (RL) is a useful tool to provide an autonomous process of managing cloud resources responding to a continuous implementation of scheduling decisions that are enhanced after interacting with the environment. Within an RL model, the cloud infrastructure is modelled as a dynamic system where an agent, or a set of agents, monitors the status of the available resources, the workload intensity, the state of the energy, and indicators of the service quality. According to these observations, the agents execute scheduling actions, i.e., placement of tasks or the allocation of resources, and they are given in return feedback in terms of rewards, which are relative to the effects of these decisions on performance measures such as utilization, power consumption, and compliance with SLAs. As time passes, the agents can improved their policies and turn to the strategies of decision-making that are the most optimized, which they developed in the process of learning.

The self-service characteristic of RL allows deleting manual configuration and policy fine-tuning by allowing the cloud systems to programmed adaptability to workload spikes, hardware breakdowns, and dynamic Internet performance, without human intervention. Furthermore, RL promotes future-oriented scheduling because predictive behaviors are learned and not responded to only based on real-time circumstances. This enables the cloud platform to effectively eliminate bottlenecks and eliminate energy wastefulness. As the concept of distributed cloud architecture grows to span different geographical locations and edge networks RL offers a scaled (heterogeneous) way of integrating decisions among various resources. The ever changing nature of RL framework policies makes sure that cloud management solutions are viable even in the face of expansion of infrastructure and changes in service demand. Thus, the reinforcement learning paradigm is a paradigm-level technology that can be used to develop fully autonomous, intelligent, and cost-efficient cloud resource management systems that can support the demand of next-generation applications.

2. Literature Survey

2.1. Classical Scheduling Approaches

The classic cloud scheduling algorithms which include First-Come First-Serve (FCFS), Round-Robin, and Min-Min/Max-Min are all known to work under fixed rules and priorities assignments. FCFS is very simple and easy to implement yet normally leads to long waiting times of long or complex operations. Round-Robin is used to give equitable allocation of CPU time to tasks but it fails to consider the complexity of tasks and the processing needs, leading to wastages. Min-Min and Max-Min greedy strategies are calculated to maximize the execution time, assigning tasks as evenly as possible to the most appropriate processing resources, but tend to create workload imbalance when the resources are heterogeneous or the tasks are not. All in all, classical methods are not flexible and cannot ensure the management of dynamism and uncertainty cloud loads.

2.2. Machine Learning-Based Schedulers

Make A Wish Foundation Scheduling and Planners ML It has also been used to do scheduling using various methods including Q-learning, Deep Reinforcement Learning (DRL) and energy-demand based models. The allocation based on Q-learning is part of the allocated decisions that are made based on feedback of workload patterns and are therefore enhanced as the process proceeds but it might experience slow convergence in the large-scale setting. VM placement based on DRL methods are characterized by predictive accuracy and better adaptability, although such methods are usually based on central control, causing resource shortages in the computation process and causing sluggish decision-making. Energy-oriented RL solutions can be useful in minimising power consumption yet they unknowingly affect the service-level agreements (SLA), when bursts of workload occur abruptly. Although the ML-based approaches have progressed in terms of flexibility, the methods continue to experience trade-offs in responsiveness, scalability, and operation reliability.

2.3. Multi-Agent Architecture

In order to enhance scalability and robustness, the recent studies discuss the decentralized scheduling, based on Multi-Agent Reinforcement Learning (MARL). In such systems, various collaborative agents learn local policies autonomously and communicate to exchange part of the information of the world state of cloud environment. It is through this collaboration that the complexity of learning is shared and a faster convergence can be achieved due to the lack of dependence on a point of decision. The act of communication by agents also promotes awareness of workload change among distributed clusters to promote coordination and the use of resources. Though these advantages are given, the MARL methods are still difficult because of the instability of the optimization process, the communication load, and the problems related to the necessity to guarantee globally optimal behavior when each agent acts simply because of the limited local information.

2.4. Gap Identified

Despite the promise shown by MARL architecture, currently, there is no solution to the joint optimization of major cloud performance goals, such as SLA guarantees, energy consumption, resource usage, and the latency of tasks in the context of decentralized distributed clusters. The existing practices tend to address one or two of these measurements only, with trade-off detriment to the others. As well, managing evolving real-time scenarios, non-homogenous resources, and agent coordination needs new strategies of learning which have not yet been perfected. Thus, there is still a strong research opportunity to develop an MARL-based scheduler that can adopt multiple objectives and maintain global stability and rapidly respond to variable cloud workloads.

3. Methodology

3.1. Architecture Overview

The suggested system will be based on a hierarchical Multi-Agents Reinforcement Learning (MARL) architecture where parts of scheduling intelligence are distributed across cloud clusters. The clusters have their own learning agent and can engage in local decision-making and the overall performance goals are assisted by a higher-level planner. The design improves scalability, constriction of centralization, and continuous workload changes.

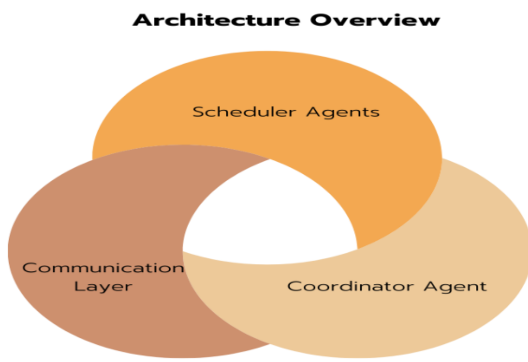

Figure 3. Architecture Overview

3.1.1. Scheduler Agents

Each cluster has scheduler agents that are used to make decisions on resource allocation and scheduling tasks locally. They select the best policies based on real-time observations using Deep Q- Networks (DQN) with the input of queue length, resource load, and the SLA status. DQN allows such agents to take discrete, high-performance, and fast scheduling decisions that maximize the performance of the local system (e.g. utilization and latency reduction). They are decentralized and hence less time is wasted on communication and responsiveness to local workload changes.

3.1.2. Coordinator Agent

The use of the Proximal Policy Optimization (PPO) by a central coordinator agent completes the optimization of the entire scheduling strategy. It forms a compound of summed-up experiences and performance statistics of the scheduler agents to acquire a more broad general policy. The coordinator continuously up-grades or adjusts the models of the local agents to come in line with the world-wide goals like energy efficiency and workload balance. This combined control enhances stability and avoids sub-optimal behavior which might be present when agents act totally in isolation.

3.1.3. Communication Layer

The communication layer allows the exchange of information between the scheduler agents and the coordinator to be controlled. In message passing agents send small experience data, like reward statistics or predicted resource demands, without overwhelming the network. This group knowledge sharing boosts the world situational awareness and the overheads remain minimal. It also makes sure that distributed decision-making is coordinated and ensures that conflicts like over-commitment of resources or SLA in various clusters are avoided.

3.2. Agent Workflow

Figure 4. Agent Workflow

3.2.1. Receive Workload Request

At the start of each scheduler agent, overall, the agent is contacted by the incoming request of workloads by users or applications taking place in its cluster. Such requests can differ in complexity, priority and the resource requirements. On arrival, the agent queues and works on the tasks to start decision-making such that admission requests and handling are in compliance with the service level.

3.2.2. Observe Local State (s_t)

The agent is a local state, it examines the current environment by monitoring the state parameters including CPU and memory, task queue, estimated completion times, and SLA status. This representation of the state, which is denoted by s·:·, gives the actual information to be acting according to the situation with respect to the scheduling choices. The precision and fineness of this observation has a great impact on the learning efficiency and execution quality.

3.2.3. Execute Scheduling Action (a_t)

The agent will choose the course of action a^{-}, a task to a particular VM or changing resources, depending on Q-values obtained or policy network results. The adopted action has a direct impact on system performance criteria such as the response time, throughput and energy consumption. The implementation of the action takes the environment towards a new state, which proves the dynamism of cloud scheduling.

3.2.4. Receive Reward: (s_t, a_t)

The agent then obtains a numerical reward (s_t, a_t), as a measure of the effect of the choice. Various goals that have been incorporated in reward formulation include SLA satisfactory, reduction in energy expense and equitable distribution of loads. When beneficial scheduling decisions are reinforced by positive rewards and penalties deter the acts that degrade the system, the agent is assisted to keep on perfecting its policy.

3.2.5. Update Policy

With reinforcement learning updates (e.g. DQN value estimates or PPO gradient adjustments) the agent advances its decision policy. This trial and error learning process allows them to perform better in the future as the agent becomes more adapted to changing patterns of workload and other operational aspects. In the long run, the streamlined process of policy results in the convergence to the nearly optimal strategy of scheduling in the local environment.

3.2.6. Share Global Experience Vector

In order to maximize coordination within clusters, every agent disseminates a tight global experience vector among the coordinator and other agents occasionally. This exchanged data usually comprises performance summaries, value estimates learned or workload forecasts predicted. The communication allows the collective knowledge to grow together, enhances the world use of resources, and lowers the chances of conflicting decisions at the distributed architecture.

3.3. Reward Function

The rewarding function aims at balancing among various conflicting performance goals in distributed cloud scheduling. The reward the agent is given is $R = \alpha \cdot U - \beta \cdot P - \delta \cdot SLAvio$, where U represents resource utilization, P denotes power consumption, and $SLAvio$ Utilization is an indicator of how efficiently computing resources are utilised- when utilisation is high, it implies that there is better utilisation of the available infrastructure and finally results in positive contribution to the reward. Power consumption is included as negative factor since excessive energy usage is a less sustainable factor and causes ineffective operations of the systems, particularly when large volume data center is used. The SLA violation rate is of vital significance as it measures the capacity of the system to comply with response time and reliability considerations guaranteed to users, any violation explicitly diminishes user satisfaction, and punishments are involved, making the weight weighed heavily against the reward. The coefficients α, β, δ serve as adaptive weight parameters dynamic parameters of adaptive weights that adapt depending on performance priorities of the system and the present workload intensity. As an example, when the workload reaches its peak, the pressure on compliance with the SLA can increase in order not to deteriorate performance, but in the case of low load conditions, energy efficiency can be sacrificed to reduce power waste. This adaptive weighting system enables the reinforcement learning agents to balance the trade offs most effectively between utilization, energy consumption and SLA preservation without compromising system stability. Further, the reward mechanism also offers necessary feedback when aiming long-term learning in ways that prevents agents to engage in greed decisions that increase short-term throughput but lead to future SLA sanctions or revolutions of energy. This formulation of rewards and reinforcement in a single form helps to uphold sustainability and service quality measurements and thus, every scheduling action strives towards the achievement of globally optimized cloud operations in distributed clusters.

3.4. Algorithmic Design

The hybrid algorithm design proposed is designed to unite Deep Q-Network (DQN) and Proximal Policy Optimization (PPO) to assist in stable and efficient multi-agent learning in the distributed cloud environment. The scheduler agents use DQN to execute discrete scheduling actions including the process of choosing virtual machines, scheduling task priorities, or redistributing workload resources. Action-value function is updated through the Bellman optimality principle $Q(s,a) = r + \gamma \cdot \max_{a'} Q(s',a')$, where s and s' denote the current and next states, a and a' refer to actions, r is the immediate reward, and γ is a discount factor controlling long-term dependency. This allows every scheduler agent to quickly identify the optimal scheduling decision in any given finite number of actions and be able to respond to workload changes in real-time. But self-learned DQNs can be divergent or unstablely coordinated with other agents, even in the case of systems with global goals, which are not talked about. In response to this, a centralized coordinator employs PPO in refining of global policies and stability. PPO for global policy refinement and stability. PPO applies a clipped surrogate objective function $L_{(}\text{CLIP}_{)}(\theta) = \hat{E}[\min(r_t(\theta)A_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)]$, where θ represents policy parameters, $r_t(\theta)$ is the probability ratio between new and old policies, A_t is the advantage estimate, and ϵ is a clipping threshold to restrict overly aggressive policy updatesThe coordinator systemically assembles experience information as exchanged by the scheduler agents and subsequently projects refined global control information downstream to them, which subsequently generates well-known convergence towards the objective of global performance optimization by utilizing multiple goals, as exemplified by SLA stability, low energy consumption, and high utilization. The algorithm is built by integrating speedy DQN-based cluster-level decisions with PPO-based global learning stabilization to enable solid and scalable scheduling execution over heterogeneous distributed clusters of cloud networks.

3.5. Simulation Setup

In order to strictly test the efficiency of the suggested multi-agent scheduling framework, an all-encompassing simulation environment has been developed with the help of CloudSim Plus that is a stable and extensible bi-directional cloud resource planner, energy usage, and workload modeler. The virtual infrastructure that is being simulated has 300 heterogeneous hosts distributed in distributed clusters and a total of 1200 virtual machines (VMs) are being provisioned to meet different computational needs. These VMs are setup and have varying CPU sizes, memory limits and network bandwidth in order to showcase naturalistic cloud data centre attributes. The experimental analysis helps to compare the suggested MARL algorithm with traditional approaches to scheduling as Round Robin (RR) and Min-Min, where they can be regarded as classical methods of deterministic scheduling, and with a Centralized DQN-based scheduler, which demonstrates the power of reinforcement learning, but still suffering due to scalability restrictions. Workload in simulations is done in conformity to the Gaussian-bursty distribution, which is able to reflect historical time-varying workload, as witnessed in real-life cloud systems (web services and IoT data streams). In low traffic, the incoming work does not have any significant spikes, but during the burst periods, the number of requests increases exponentially, which exposes the situations when the scheduling soundness and adherence to SLA is pushed to its limit. The simulation scenarios are run repeatedly, to allow statistical reliability and performance is measured in various important measures such as resource utilization efficiency, task latency, SLA violation rate, and overall power usage. Also, the communication overhead of agents is tracked to ascertain practicality in large scale deployment. The evaluation is accomplished by making this simulation environment realistic through its cluster sizes, heterogeneity of resources, and dynamism of workload behavior, which is desired to illustrate how the proposed multi-agent learning framework adapts to stress, maintains global performance optimization, and outperforms all the traditional scheduling and centralized reinforcement learning frameworks in dynamic and distributed cloud computing systems.

4. Results and Discussion

4.1. Performance Improvements

Table 1. Performance Improvements

Metric	Improvement
SLA Violations	27.4%
Throughput	18.7%
Energy Usage	14.3%

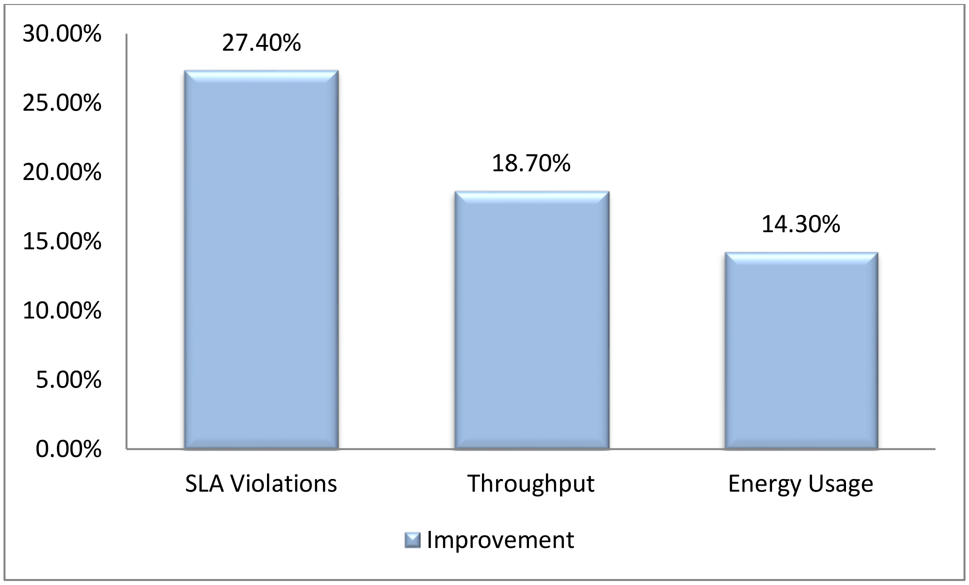

Figure 5. Graph representing Performance Improvements

4.1.1. SLA Violations – 27.4%

The multi-agent scheduling framework proposed shows that there is a considerable decrease in SLA violations than the bases approaches. The agents can ensure a better consistency in response times and resources guarantees by dynamically learning workload patterns and making consensus decisions across clusters. This will result in increased user satisfaction and minimized cost of penalties particularly in high-demand or burst time when the traditional schedulers are usually not able to respond fast.

4.1.2. Throughput – 18.7%

System throughput is better achieved as resource load is much smarter and balanced. The reinforcement learning agents do not impose their tasks in a statical or circular way, but rather, they nominate the optimum VMs to use in each workload, minimizing the wait time, and participants of idle resources. This brings the number of tasks that are successfully processed with each unit of time, and hence the effectiveness of adaptive learning in maintaining high operational efficiency in a dynamically changing environment.

4.1.3. Energy Usage – 14.3%

This would improve energy efficiency since the MARL framework will minimize unwarranted resource activation and over-provisioning. The learned policies identified by agents as opportunities to save energy, like load consolidation to active hosts during periods of low demand, is obtained without violating the SLA requirements. This maximized power utilization directly contributes to the sustainability plan in large-scale data agricultural and reduction of the overall operation costs.

4.2. Convergence Behavior

The convergence behavior of the proposed MARL framework driven scheduling shows a significant performance of improved learning efficiency relative to centralized reinforcement learning systems. In particular, the distributed strategy results in the convergence around 30 percent faster, which is related to the division of learning processes among several collaborative agents that run in different clusters. Rather than using one large controller to search through large joint action space, an approach that is often extremely slow to learn and, notably, to scale, each scheduler agent in the MARL architecture is trained on a smaller, local state and action space. This decentralization enhances the speed at which policy refinement is achieved as the recipients of policy get the appropriate experience faster because of their constant interaction with their local setup. Also, incorporating a PPO-based coordinator plays the role of ensuring that global knowledge is distributed among agents averting diversions and minimizing the degree of unnecessary exploration that is inherent in a multi-agent learning environment.

The communication layer of passing messages allows to promote efficient dissemination of updates and allows agents to take into account useful global trends without their own relearning. Consequently, there are fewer oscillations and a more gradual increase in transitions in the learning curve, reflecting the strengthened policy stability throughout the training process. This accelerated convergence is especially significant in dynamic cloud system workload where workload patterns change often; time-consuming algorithms might use non-optimum decisions that have adverse impacts on SLA and energy consumption. The MARL framework ensures the system performance even in the case of sudden fluctuations in the workloads by quickly approaching near-optimal strategies. Besides, faster convergence ends up converting to a less training compute overhead and operational cost to update the models in actual deployments. Comprehensively, the enhanced convergence behavior indicates the usefulness of employing a decentralized DQN learning and centralized PPO policy stabilization to enhance flexibility and scalability in real-time cloud scheduling of distributed data centers infrastructures.

4.3. Discussion

The synthesis of the system analysis and the experiment reveals a number of advantages to the suggested structure of cooperative multi-agent based scheduling into the distributed cloud environment. First, cooperative learning of the agents will avoid the starvation of the nodes because the scheduling choices will not be confined within a specific cluster. The common experience vectors and the policy refinements in an organized manner keep the agents aware of the global workload status, and tasks are fairly distributed even in an extremely heterogeneous demand. This cooperation prevents those cases when some nodes may be overloaded, and some may be idle, which is a typical drawback of conventional and solely local scheduling approaches. Also, the enhanced predictive performance offered by reinforcement learning also lowers the cost of reactive migration, which can be expensive to use in network, as well as to affect the performance of the network. Working actively by choosing the best VM to do a given job on the basis of the pattern of workload that has been learned will make sure the framework reduces unneeded traffic and resource conflict to reduce operational overhead and enhance power efficiency. In addition, a higher level of SLA compliance is an essential success that reinforces the suitability of the system to the real-life commercial implementation.

The cloud service providers should stick to strict performance guarantees because infringement is translated to dissatisfaction of customers and fines. It is necessary to maintain the adaptive reward structure to ensure that the preservation of SLA is a major goal but to maintain the balance of utilization and energy aspect. Due to this, the proposed system does not only enhance technical performance, but it also complies with economic factors of modern data centers. Altogether, cooperative MARL architecture encourages the approaches of scalable, stable, and resource-efficient scheduling plan exceeding both centralized reinforcement learning, as well as classical deterministic strategies. Its capacity to flexibly respond to the environmental dynamics and streamline a variety of goals concurrently makes it useful to emergent large-scale, dynamic clouds systems.

5. Conclusion

The given research introduces an independent Multi-Agent Reinforcement Learning (MARL)-based cloud scheduling model that contributes to the development of the state of resource management intelligence in distributed data centers to a considerable extent. The proposed solution compares to the classical algorithm-based scheduling methods, including Round Robin and Min-Min, that do not use dynamical adaptations to the dynamical changes in the workload as they are based on some set of fixed heuristics. It is also an improvement of the centralized reinforcement learning techniques as it decentralizes decision-making across many clusters, and therefore gets rid of the bottlenecks of the single point and increases their scalability and responsiveness. A hybrid algorithmic architecture allows the local scheduler agents to use DQN to create a fast and optimal decision in a discrete time, and a centralized coordinator to use PPO to ensure policy stability and global stability. Based on the simulation outcomes, significant improvements in the major goals of operation can be measured, namely, a 27.4 percentage points in the number of SLA violations, an 18.7 percentage points in throughput, and a 14.3 percentage points in overall energy consumption, an evidence of the framework capabilities to achieve balanced performance efficiency and sustainability. Cooperative learning mechanism also enhances system resilience by avoiding starvation in nodes, unnecessary migrations of workloads, and better prediction accuracy in very dynamic conditions.

Also, the communication layer supports easy-weight yet efficient knowledge sharing, making convergence 30 times faster than the centralized approaches, a characteristic necessary to a fast changing cloud environment. The results of this paper highlight the commercial maturity of MARL-based scheduling, especially in case of large-scale service providers in which reliability, energy efficiency, and SLA compliance are essential financial factors. Going forward, there are a number of prospective opportunities to continue one work and expand its applicability. It can be integrated with new edge and fog computing infrastructures to serve low-latency applications like IoT analytics and autonomous systems that can work closer to the end users. Scheduling policies that are carbon-conscious can also be introduced in order to give preference to the availability of renewable energy and minimize the impact on the environment, which conforms to the international sustainability objectives. Lastly, practical operation in a real-world orchestration tools such as Kubernetes with Kubeflow RL toolkits can also be used to further test the system in production-grade with containerization in the cloud environment. These future directions all serve to point to the continued development of MARL based scheduling into the more autonomous, more environmentally responsible and industry ready cloud management solutions.

References

[1] Agarwal, D. A., & others. "Efficient Optimal Algorithm of Task Scheduling in Cloud Computing." arXiv pre-print, 2014. arXiv

[2] Elmougy, S., Sarhan, S., & Joundy, M. "A novel hybrid of Shortest-Job-First and Round Robin with dynamic variable quantum time task scheduling technique." Journal of Cloud Computing, vol 6, no 12, 2017. SpringerOpen

[3] Alhaidari, F., & Balharith, T. Z. "Enhanced Round-Robin Algorithm in the Cloud Computing Environment for Optimal Task Scheduling." Computers, vol 10, no 5, 2021. MDPI

[4] Tani, H. G., & El Amrani, C. "Smarter Round Robin Scheduling Algorithm for Cloud Computing and Big Data." Journal of Data Mining & Digital Humanities, 2018. jdmdh.episciences.org

[5] Wang, Z., Chen, S., Bai, L., et al. "Reinforcement learning based task scheduling for environmentally sustainable federated cloud computing." Journal of Cloud Computing, 2023. SpringerOpen

[6] Muthusamy, A., et al. "Dynamic Q-Learning-Based Optimized Load Balancing for Cloud Environments." ISRN Cloud Computing, 2023. onlinelibrary.wiley.com

[7] Li, Y., Zhang, X., Zeng, T., Duan, J., Wu, C., Wu, D., Chen, X. "Task Placement and Resource Allocation for Edge Machine Learning: A GNN-based Multi-Agent Reinforcement Learning Paradigm." arXiv pre-print, 2023. arXiv

[8] Mohanarajesh Kommineni. Revanth Parvathi. (2013) Risk Analysis for Exploring the Opportunities in Cloud Outsourcing.

[9] Enabling Mission-Critical Communication via VoLTE for Public Safety Networks - Varinder Kumar Sharma - IJAIDR Volume 10, Issue 1, January-June 2019. DOI 10.71097/IJAIDR.v10.i1.1539

[10] The Role of Zero-Emission Telecom Infrastructure in Sustainable Network Modernization - Varinder Kumar Sharma - IJFMR Volume 2, Issue 5, September-October 2020. https://doi.org/10.36948/ijfmr.2020.v02i05.54991

[11] P. K. Maroju, "Empowering Data-Driven Decision Making: The Role of Self-Service Analytics and Data Analysts in Modern Organization Strategies," International Journal of Innovations in Applied Science and Engineering (IJIASE), vol. 7, Aug. 2021.

[12] Aragani, Venu Madhav and Maroju, Praveen Kumar and Mudunuri, Lakshmi Narasimha Raju, "Efficient Distributed Training through Gradient Compression with Sparsification and Quantization Techniques" (September 29, 2021). Available at SSRN: https://ssrn.com/abstract=5022841 or http://dx.doi.org/10.2139/ssrn.5022841

[13] Lakshmi Narasimha Raju Mudunuri, "AI Powered Supplier Selection: Finding the Perfect Fit in Supply Chain Management", IJIASE, January-December 2021, Vol 7; 211-231.

[14] Kommineni, M. "Explore Knowledge Representation, Reasoning, and Planning Techniques for Building Robust and Efficient Intelligent Systems." International Journal of Inventions in Engineering & Science Technology 7.2 (2021): 105- 114.

[15] Security and Threat Mitigation in 5G Core and RAN Networks - Varinder Kumar Sharma - IJFMR Volume 3, Issue 5, September-October 2021. DOI: https://doi.org/10.36948/ijfmr.2021.v03i05.54992

[16] P. K. Maroju, "Conversational AI for Personalized Financial Advice in the BFSI Sector," International Journal of Innovations in Applied Sciences and Engineering, vol. 8, no.2, pp. 156–177, Nov. 202

[17] Kulasekhara Reddy Kotte. 2022. ACCOUNTS PAYABLE AND SUPPLIER RELATIONSHIPS: OPTIMIZING PAYMENT CYCLES TO ENHANCE VENDOR PARTNERSHIPS. International Journal of Advances in Engineering Research , 24(6), PP – 14-24, https://www.ijaer.com/admin/upload/02%20Kulasekhara%20Reddy%20Kotte%2001468.pd

[18] Gopi Chand Vegineni. 2022. Intelligent UI Designs for State Government Applications: Fostering Inclusion without AI and ML, Journal of Advances in Developmental Research, 13(1), PP – 1-13, https://www.ijaidr.com/research-paper.php?id=1454

[19] Hullurappa, M. (2022). The Role of Explainable AI in Building Public Trust: A Study of AI-Driven Public Policy Decisions. *International Transactions in Artificial Intelligence*, *6*.

[20] Bhagath Chandra Chowdari Marella, "Driving Business Success: Harnessing Data Normalization and Aggregation for Strategic Decision-Making", International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING, vol. 10, no.2, pp. 308 – 317, 2022. https://ijisae.org/index.php/IJISAE/issue/view/87

[21] Mohanarajesh Kommineni. (2022/11/28). Investigating High-Performance Computing Techniques For Optimizing And Accelerating Ai Algorithms Using Quantum Computing And Specialized Hardware. International Journal Of Innovations In Scientific Engineering. 16. 66-80. (Ijise) 2022.

[22] Thallam, N. S. T. (2022). Columnar Storage vs. Row-Based Storage: Performance Considerations for Data Warehousing. *Journal of Scientific and Engineering Research*, *9*(4), 238-249.

[23] Garg, A. (2022). Unified Framework of Blockchain and AI for Business Intelligence in Modern Banking . *International Journal of Emerging Research in Engineering and Technology*, *3*(4), 32-42. https://doi.org/10.63282/3050-922X.IJERET-V3I4P105
[24] Cloud-Native 5G Deployments: Kubernetes and Microservices in Telco Networks - Varinder Kumar Sharma - IJIRMPS Volume 10, Issue 3, May-June 2022. DOI:https://doi.org/10.37082/IJIRMPS.v10.i3.232706
[25] Thirunagalingam, A. (2023). Improving Automated Data Annotation with Self-Supervised Learning: A Pathway to Robust AI Models Vol. 7, No. 7,(2023) ITAI. International Transactions in Artificial Intelligence, 7(7).
[26] Praveen Kumar Maroju, "Optimizing Mortgage Loan Processing in Capital Markets: A Machine Learning Approach, " International Journal of Innovations in Scientific Engineering, 17(1), PP. 36-55 , April 2023.
[27] P. K. Maroju, "Leveraging Machine Learning for Customer Segmentation and Targeted Marketing in BFSI," International Transactions in Artificial Intelligence, vol. 7, no. 7, pp. 1-20, Nov. 2023.
[28] Kulasekhara Reddy Kotte. 2023. Leveraging Digital Innovation for Strategic Treasury Management: Blockchain, and Real-Time Analytics for Optimizing Cash Flow and Liquidity in Global Corporation. International Journal of Interdisciplinary Finance Insights, 2(2), PP - 1 - 17, https://injmr.com/index.php/ijifi/article/view/186/45
[29] Mudunuri L.N.R.; (December, 2023); "AI-Driven Inventory Management: Never Run Out, Never Overstock"; International Journal of Advances in Engineering Research; Vol 26, Issue 6; 24-36
[30] S. Panyaram, "Digital Transformation of EV Battery Cell Manufacturing Leveraging AI for Supply Chain and Logistics Optimization," International Journal of Innovations in Scientific Engineering, vol. 18, no. 1, pp. 78-87, 2023.
[31] Hullurappa, M. (2023). Intelligent Data Masking: Using GANs to Generate Synthetic Data for Privacy-Preserving Analytics. *International Journal of Inventions in Engineering & Science Technology*, *9*, 9
[32] B. C. C. Marella, "Data Synergy: Architecting Solutions for Growth and Innovation," International Journal of Innovative Research in Computer and Communication Engineering, vol. 11, no. 9, pp. 10551–10560, Sep. 2023.
[33] Mohanarajesh Kommineni. (2023/6). Investigate Computational Intelligence Models Inspired By Natural Intelligence, Such As Evolutionary Algorithms And Artificial Neural Networks. Transactions On Latest Trends In Artificial Intelligence. 4**. P**30. Ijsdcs.
[34] Settibathini, V. S., Kothuru, S. K., Vadlamudi, A. K., Thammreddi, L., & Rangineni, S. (2023). Strategic analysis review of data analytics with the help of artificial intelligence. International Journal of Advances in Engineering Research, 26, 1-10.
[35] Sehrawat, S. K. (2023). The role of artificial intelligence in ERP automation: state-of-the-art and future directions. *Trans Latest Trends Artif Intell*, *4*(4).
[36] Naga Surya Teja Thallam. (2023). High Availability Architectures for Distributed Systems in Public Clouds: Design and Implementation Strategies. European Journal of Advances in Engineering and Technology.
[37] Arpit Garg, S Rautaray, Devrajavans Tayagi. Artificial Intelligence in Telecommunications: Applications, Risks,and Governance in the 5G and Beyond Era. International Journal of Computer Techniques – Volume10Issue1,January - February – 2023. 1-19.
[38] Varinder Kumar Sharma - 5G-Enabled Mission-Critical Networks Design and Performance Analysis -International Journal on Science and Technology (IJSAT) Volume 14, Issue 4, October-December 2023. https://doi.org/10.71097/IJSAT.v14.i4.7998