FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 1, January, 2026, pp 291 – 295
DOI: https://doi.org/10.33003/fjs-2026-1001-4378

AUTONOMOUS QOS-AWARE RESOURCE PROVISIONING IN DISTRIBUTED CLOUD ENVIRONMENTS: A UNIFIED PERSPECTIVE FROM ROBOTICS, EDGE, AND BPM DOMAINS

***[1]Muhammad Tella, [1]Kabiru Ibrahim Musa, [1]Mahmud Ahmed Usman and [2]Fatima Umar Zambuk**

[1]Department of Management and Information Technology, Faculty of Management Sciences, Abubakar Tafawa Balewa University, Bauchi State, Nigeria.
[2]Department of Computer Science, Faculty of Computing, Abubakar Tafawa Balewa University, Bauchi State, Nigeria.

*Corresponding authors' email: tmuhammad@atbu.edu.ng

ABSTRACT

Distributed computing has been a long existing technology that has allowed computer collaboration in terms of handling complex tasks with very high efficiency. The trend in distributed computing has witnessed some revolution in terms of the domains that use it as well as the methodologies in its implementation. Recently, some among the range of domains that utilizes distributed computing include robotics, mobile edge networks and business process management (BPM). Each domain may have a different use case but they commonly leverage on the increasingly growing intelligence in terms of resource provisioning. Some of the challenges systems using distributed computing have attempted to solve includes dealing with dynamic workloads while at the same time meeting certain constraints like Quality of Service (QoS) and Service Level Agreement (SLA). Traditionally robotics, mobile edge networks and BPM as distinct domains have differing use cases and cannot be autonomously utilized by systems that may require the use of all three domains. This research work focuses on synthesizing the potentials of the three key research areas, namely multi-agent cloud robotics, heuristic and learning-based edge computing, and BPM with the aim of obtaining an efficient resource allocation in a dynamic workload environment. The proposed framework combines the MAPE-K loop, EDSAE and MOTCO techniques through enhanced K-Means clustering. Comprehensive experiment demonstrates that after a 24-hour simulation using Extended Container CloudSim and BPM workload traces revealed that the proposed framework outperformed both static provisioning and reactive auto-scaling strategies. The proposed framework shows significant reduction in response time, energy consumption and execution costs. Specifically, SLA violation is reduced by 2.3%, a boost in CPU utilization of 89.1% and improved throughput of 14.6 tasks/sec.

Keywords: Artificial Intelligence (AI), Edge Computing, Quality of Service (Qos), Machine Learning (ML), Resource Provisioning

INTRODUCTION

Due to the fast development of computer technology this process has accelerated the synergy of cloud and edge computing with robotics. Each of the three areas have a distinct advantage, yet the melding of these technologies appears to erase the separation between them. The scalability of resource from cloud computing, the close proximity of source data and processing point which reduces latency on the part of edge computing and the rapid facilitation of autonomous interaction with the physical world offered by robotics imply that fusion of these three will be significant and necessary for realizing components of cyber-physical systems (CPS) such as autonomous vehicles, precision agriculture, remote health care, and smart manufacturing.

In a survey by Afrin et al. In (2021) they recognised more than 10 challenges relating to some operations on the cloud such as resource allocation, task scheduling and offloading. These challenges were predominantly towards optimal resource allocations in multi-agent cloud robotics environments that are compute-sensitive and are also latency sensitive. They found the implied latency problems with conventional cloud systems to be among the challenges the authors. They claimed that the very latency embedded in those systems render them impractical for responsive robotic applications.

Edge computing can easily solve latency issues by bringing the processing closer to the data source to enable quicker and more reliable responses (Garg et al.).

The final efficiency of this synergy is due to the fact that the features of cloud computing, robotics and edge computing can be combined because the tasks can be offloaded based on latency, energy and QoS requirements. If we combine the benefits of these three technologies, we can make the businesses better. It is because the businesses can gain enhanced competitive edge, by providing on-demand services for accessing a common pool of shared virtual resources such as VMs and cloud containers.

Moreover, recent advancements in Artificial Intelligence (AI) and Machine Learning (ML) techniques can enhance task scheduling, as well as predicting the outcomes of operations. Hence, real-time workload characteristics and network behavior changes are more than manipulations over which control can equally be augmented (Garg et al., 2021)

Despite these developments however, there does not appear to be an integrated approach to resource provisioning. Cloud computing typically represents intense dependency on the centralized processing offloading while edge computing is mainly concerned with service placement and replication. Also, BPM in multi-cloud environment center lean too much onto container auto-scaling under SLAs. Generally, the problem is still the ability of dynamic resource allocation while ensuring Qos which is still endemic, but solutions are specific to the domain.

If care delivery of different services is not well unified the opportunity to forge partnerships is missed, and inefficiency can become a norm. In addition, BPM systems seamlessly do not exploit innovations like RL-based placement in edge computing, and robotic frameworks hardly get exposure in enterprise clouds. Effectively what this means is that there is no cross-domain which makes creation of scalable and resilient provisioning frameworks hard.

One provisioning model is required for these issues to be fixed. A cross-domain model should accommodate edge,

robots, or cloud handover solutions using heuristic, learning-based and agent-driven strategies and allow QoS-aware resource allocation across boundaries. Such a model will balance efficiency with impermeability and enable intelligent orchestration in complex distributed environments. Further emphasis was made by recent advances in intelligent cybersecurity systems regarding the importance of adaptive, QoS-aware decision-making in distributed environments. For instance, in Aji et al. (2025), it was demonstrated that Bayesian-optimized ensemble learning significantly improves system reliability and performance by dynamically tuning models to evolving threat conditions, thereby achieving near-perfect accuracy and robustness in real-time phishing detection. The value of automatic, learning-driven frameworks for managing complexity, performance, and uncertainty is enforced by these findings.

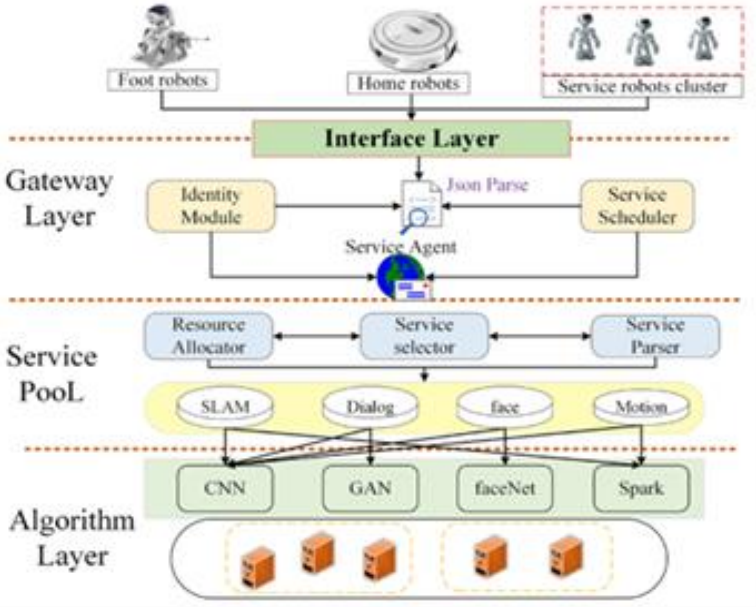

Figure 1: Provisioning in Distributed Cloud Environments

Objectives and Motivation

Right now, it seems fair to say that research edge computing, robotics and business process management (BPM) are in a kind of disarray with each field growing primarily on its own terms. The main principle of edge computing is to process the data as close to the source as possible which of course aids us to alleviate bandwidth usage and latency but is also crucial in real-time applications. Most robotics research focuses on mechanical design, control, and autonomy. The focus of robotics research is to create intelligent machines that can move, manipulate objects, and perform complex actions. On the other hand, BPM is primarily focused on enhanced business process efficiencies and enhanced efficiencies and agile flexibility through automation of enterprise workflows. While these fields are not completely independent, their trajectories have not been studied together. Because there has not been this examination we are left to build systems and we are trapped between the two worlds not knowing how best to combine their strengths.

The splitting of these strengths into these fields creates barriers, particularly for industries that are looking to migrate towards ever more connected and intelligent systems. The process can be illustrated as follows, for instance enrichment of Business Process Management (BPM) with Internet off Things (IoT) features to streamline real-time decision making process and flexibility. But we can't do this, because there is not a standard protocol, and our technology is not compatible. Likewise, in the sense edge computing answers is a way to process data on the field from robotic systems, this heterogeneity stands in the way for robots to be embedded in BPM systems. The differences in focus area and approach across these fields restrict design capability to produce an integrated, adaptive and intelligent systems but also results in a waste of resources.

A possible answer to them is a targeted initiative to the development of clear frameworks and standards that allow for integration. Novel concepts such as Internet of Robotic Things (IoRD), need to fill these gaps through the combination of Robotics, edge computing, and BPM into unified systems that operate in real-time with specifically meaningful data and autonomously make new decisions and perform particular actions. It integrates into the system to work with the operations which leads to improved operational efficiency, agility, and complex tasks handling.

Related Work

With the constant development of cloud computing, cyber-physical systems, and cybersecurity infrastructure, there is a demand for intelligent, flexible, and automated solutions for resource management, edge computing, and scenario generation. Many areas have experienced substantial progress ensuring the foundation needed for creating robust, scalable, and efficient digital infrastructures.

Costa et al. (2020) proposed a cognitive enabling framework to automate definition and deployment of cyber range scenarios based on the Virtual Scenario Description Language (VSDL). By automatically adapting the virtual infrastructures createdcycle to tackle the scalability problems of the typical cyber ranges, their research focus on solving these issues. They use satisfiability modulo theories (SMT) for verifying correctness of scenarios. They highlight the necessity of verifiability and the ability to compose systems in creating models and interfaces for cyber infrastructures, which strengthens the foundation needed to build programmable and flexible cyber training environments.

Meanwhile, Afrin et al. (2021), in a survey on multi-agent cloud robotic systems highlighted the key point that using MMOS can enhance the autonomy of robot swarms by enabling teamwork and sharing of resources and the hybridization with edge and cloud infrastructures. It also provides a classification system for resource allocation strategies.

Specifically, such distributed environments face issues such as latency sensitivity, energy constraints, and dynamic coordination in heterogeneous multi-level computing frameworks.

Saif et al. Multi-agent framework for resource management and orchestration in containerized multi-cloud environments (2022) Through deep learning, they predict the workloads and by swarm intelligence they assign the resources in their framework. This method guarantees QoS and addresses problems of both over- and under-provisioning. This research is particularly relevant in the context of elastic Business Process Management (BPM) systems, where flexible provisioning policies are needed to cope with changing demands.

Lastly, Liu et al. (2024) Lazy-Encoder (Omnes et al., 2024) proposed an image codec based on INRs that allows this exact kind of simple decoding. This overcomes the drawbacks of existing deep learning-based image compression for edge devices. They contribute to boost the whole ecosystem around edge computing with better speed and quality. At their heart, they employ a mixed autoregressive model as they seek to both improve decoder performance and reduce computation needs in a way that is automated, scalable, and takes smart decisions when it comes to resource usage. They note the further blurring of the distinction between edge and cloud technologies, AI at the core of decision making, and the need to integrate cyber-physical systems in near-real time.

MATERIALS AND METHODS

In this work, we proposed a unified multi-agent framework for QoS-aware autonomic resource provisioning. The framework integrates the advantages of robotics, edge computing, and Business Process Management (BPM) into decentralized containerized multi-cloud ecosystems. Autonomic computing with the MAPE-K loop, multi-agent systems, and AI-based prediction (EDSAE) dynamically allocates resources subjected to QoS constraints such as SLA violation rate, response time, execution cost, and energy efficiency (EDSAE). Collectively, these studies deliver foundational methods and theoretical frameworks to modern distributed systems

Workloads with Improved K-Means

We employ an improved K-Means clustering technique to combine the diverse workloads regularly encountered in the BPM and robotic workloads. The input centroid selection method uses the following parameters to classify workloads into cpu-intensive and I/O-intensive types.

Range Calculation

$$Range = \max(C(i)) - \min(C(i)) \quad (1)$$

Where $C(i)$: The workload characteristic vector for the ith instance (e.g., CPU or I/O demand).

$\max(C(i))$, $\min(C(i))$: Maximum and minimum values across the $C(i)$ vector.

Cutoff Threshold

$$Cutoff = \frac{Range}{2} \quad (2)$$

Note: The cutoff is used to differentiate between CPU-intensive and I/O-intensive workloads.

Objective Function (Error Minimization)

$$J(B) = \sum_{i=1}^{k} \sum_{a_i \in B_i} \|a_i - \mu_i\|^2 \quad (3)$$

Where

$J(B)$: Total within − cluster squared error (objective to minimize).

k: Number of clusters (typically 2: CPU and I/O intensive).

a_i : Individual workload vector assigned to cluster B_i

μ_i : Centroid (mean) of cluster B_i

$\|.\|^2$: Euclidean distance squared.

This avoids resource underutilization and eliminates noisy inputs

Workload Prediction using EDSAE

An Enhanced Deep Stacked Autoencoder (EDSAE) which enables future resource demand prediction from historical workload records. Manifold-EDSAE: A Novel Version of A(EDSAE)Traditional deep learning reconstruct hidden layer features to avoid losing information.

Used during the analyze phase of MAPE-K loop.

Extended MOTCO to Best Resource Mapping

MOTCO (Multi-Objective Termite Colony Optimization) This algorithm, known as MOTCO, maps workloads to containers or edge nodes according to the constraints of resources such as CPU, bandwidth, energy and latency in time. It optimizes for:

i. SLA violations
ii. CPU utilization
iii. Execution cost
iv. Energy efficiency

Decision-making and placement by both local and global agents is coordinated with respect to the outputs of MOTCO.

Experiment

This work assesses an agent-based autonomic methodology which integrates predictive artificial intelligent models specified within the scope of this implementation, heuristic placement of cognitive agents and autonomic container elasticity with a machine learning carrier for BPM workloads within multi cloud robotic solutions.

Setup and Metrics

*Simulator***:** Extended Container CloudSim Workload Traces: Realistic BPM workload logs have known CPU and memory demand profiles Environment: Kubernetes clusters deployed on distributed edge and cloud nodes (AWS, GCP) The experiment will have the following scenarios: 1. Static Provisioning, 2. Reactive Auto-scaling and 3. Proposed Framework.

24h of continuous simulation will be the duration of the experiment.

Here are the metrics used to derive the above metrics —

i. SLA Violation Rate (%)
ii. CPU Utilization (%)
iii. Response Time (ms)
iv. Throughput (tasks/sec)
v. Execution Cost (USD)
vi. Energy Consumption (kWh)
vii. Make span (sec)

The data shown in Table 1 illustrates the result of the experiment.

Table 1: Performance Comparison

Metric	Static Provisioning	Reactive Scaling	Proposed Framework
SLA Violation (%)	19.2	7.8	2.3
CPU Utilization (%)	54.6	73.4	89.1
Avg. Response Time (ms)	327	212	134
Execution Cost (USD)	112.7	94.3	81.5
Energy (kWh)	1340	1213	1031
Make-span (s)	139	114	96
Throughput (tps)	8.7	11.4	14.6

Here is the SLA Violation Rate vs Time graph. As shown, the proposed framework consistently maintains a lower violation rate compared to static and reactive methods.

Figure 2: A graph of SLA against Time

Figure 3: A graph of Resource Utilization by Strategy

This is the graph Resource Utilization by Strategy. As can be seen, the proposed framework shows a substantial and stable improvement in terms of the overall resource utilization, across CPU, memory, and bandwidth resources.

RESULTS AND DISCUSSION

The proposed framework was evaluated against a static provisioning and reactive auto-scaling using a 24-hour simulation in an Extended Container CloudSim environment. The results show consistent performance improvements across all evaluated metrics.

The framework shows a significant reduction of SLA violations from 7.8% and 19.2% for reactive and static strategies respectively to 2.3%, demonstrating effective proactive QoS management. There is an improvement of CPU utilization to 89.1%, indicting efficient usage of resource and reduced underutilization. There is also benefit in latency-sensitive workloads from a lower average response time of 134 ms, confirming the advantage of predictive, edge-aware provisioning.

Additionally, the proposed approach shows an achievement of lower execution cost (N112,998.12) and energy consumption (1031 kWh), reflecting improved economic and energy efficiency. There is increased system throughput of 14.6 tasks/sec, while make-span reduced to 96 sec. highlighting enhanced scheduling efficiency and scalability.

Overall, the results validate that integrating multi-agent coordination, AI-based workload prediction, and autonomic control enables superior QoS compliance and resource efficiency in distributed edge-cloud environments.

CONCLUSION

The paper proposes an intelligent multi-agent framework that integrates business process management and robotic cyber-physical systems within containerized edge-cloud architecture that is also scalable so as to enable cooperative and adaptive resource allocation. By using microwaves, the framework supports dynamic task execution across such domains as elastic business workloads and drone navigation. The paper further advances container-based cyber range platforms through the introduction of adynamic scenario engine that adapts task difficulty to trainee performance, thereby overcoming the limitations of static training scenarios. The work also embeds collaborative role-based learning into cyber range design. This feature emphasizes

teamwork and shred defense to better reflect real-world cybersecurity operations.

REFERENCES

Afrin, M., Rahman, A., Rahman, A., & Hossain, E. (2021). Resource Allocation and Service Provisioning in Multi-Agent Cloud Robotics: A Comprehensive Survey. *EEE Communications Surveys & Tutorials, 23*(2), 842-870.

Aji, L. I., Idris, I., Subairu, S. O., Noel, M. D., & Ahmed, S. (2025). Bayesian-optimized ensemble support vector machine model for phishing email detection. FUDMA Journal of Sciences, 9(12), 837–842. https://doi.org/10.33003/fjs-2025-0912-4356

Costa, G., Russo, E., & Armando, A. (2020). Automating the Generation of Cyber Range Virtual Scenarios with VSDL. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA), 13*(1), 33-52. https://doi.org/10.22667/JOWUA.2022.03.31.0033

Garg, D., Narendra, N., & Tesfatsion, S. (2021). Heuristic and reinforcement learning algorithms for dynamic service placement on mobile edge cloud. *arXiv preprint*.

Hu, S., Shi, W., & Li, G. (2022). CEC: A Containerized Edge Computing Framework for Dynamic Resource Provisioning. *IEEE Transactions on Mobile Computing, 22*(7), 3840-3854. https://doi.org/10.1109/TMC.2022.3147800

Liu, X., Chen, J., Chen, B., Liu, Z., An, B., Xia, S.-T., & Wang, Z. (2024). An Efficient Implicit Neural Representation Image Codec Based on Mixed Autoregressive Model for Low-Complexity Decoding. *arXiv preprint arXiv*, 1-10.

Saif, M. A., S. K., N., Murshed, B. A., Al-Ariki, H. D., & Abdulwahab, H. M. (2023). Multi-agent QoS-aware autonomic resource provisioning framework for elastic BPM in containerized multi-cloud environment. *Journal of Ambient Intelligence and Humanized Computing, 14*(9), 12895-12920.

Shrivastava, A., Nayak, C. K., Dilip, R., Samal, S. R., Rout, S., & Ashfaque, S. M. (2023). Automatic robotic system design and development for vertical hydroponic farming using IoT and big data analysis. *Materials Today: Proceedings*, (pp. 3546-3553.).

Valner, R., Vunder1, V., Aabloo, A., Pryor, M., & Kruusamäe, K. (2022). TeMoto: A Software Framework for Adaptive and Dependable Robotic Autonomy with Dynamic Resource Management. *IEEE Access, 10*, 51889-51907.