A Review on Autonomous and Adaptive Scheduling Framework for Next-Generation Fog Computing Systems

Abhijeet Mahapatra[1], *Ganesh Khekare*[2]
[1]Postdoctoral Researcher, Lincoln University College, Malaysia;
[2]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India
pdf.abhijeet@lincoln.edu.my, khekare.123@gmail.com

Abstract: Fog computing has emerged as an effective paradigm for modern latency-sensitive, resource-intensive and deadline-critical applications. However, for efficient scheduling of dynamic tasks and managing resources, traditional centralized approach poses significant issues. This article presents a systematic review of the literature on adaptive and autonomous scheduling algorithms in Fog computing environment. The challenges of energy efficiency, resource utilization, cost and latency have been addressed in this study. The key contributions of this work provide helpful insights for the researchers in terms of designing robust and sustainable systems that are autonomous and can adapt to the dynamic nature of incoming workload for the modern Fog-Cloud environments. By identifying key limitations, this article paves the way for future research in terms of autonomous and adaptive scheduling frameworks that are lightweight yet effective in nature.

Keywords: Autonomous and Adaptive Scheduling, Energy Efficiency, Fog-Cloud Computing, Latency, Resource Utilization, Task Offloading.

Introduction

Fog computing has emerged as an important paradigm in the bridging of the centralized Cloud system and the edge devices. Fog computing allows for the processing of information closer to its source, hence solving issues related to high latency, limited scalability, and energy inefficiency [1]. Such attributes make Fog computing indispensable for applications requiring low latency, such as healthcare monitoring, autonomous vehicles, and industrial IoT systems [2]. However, this dynamic and resource-limited nature of Fog environments contributes to the increasing demand for developing novel scheduling mechanisms which could bring better task allocation and exploitation of resources. Figure 1 shows the basic architecture of IoT-Fog-Cloud paradigm.

Next-generation Fog computing environments operate in highly dynamic and heterogeneous IoT-Fog-Cloud environments, where there is rapid fluctuation of workloads, frequent joining or failure of computational nodes, and network conditions such as latency and bandwidth vary unpredictably. Traditional centralized scheduling approaches cannot cope with these conditions because they lack the ability to make decisions in real-time variations for availability of resources, system state and application sensitivity. Many emerging applications such as healthcare, autonomous vehicles, industrial automation, AR/VR, *etc.* demand low latency and hard real-time responses, which cannot be guaranteed if tasks are offloaded to Cloud servers present at a distant place.

Moreover, Fog nodes differ widely in computational power, computational memory, computational energy capacity and hardware requirements. Hence, it is essential to have a scheduler that can autonomously assign tasks to the most suitable nodes while reducing energy consumption and latency.

As the number of IoT devices is growing at a rapid rate, centralized control has become a bottleneck and a single point of failure, reinforcing the need for decentralized, self-optimizing scheduling mechanisms. Additionally, scheduling in a Fog environment often includes addressing multiple objectives such as maximizing utilization of resources while reducing latency, energy consumption and cost. Managing these objectives simultaneously requires real-time adaptive decision making rather than fixed rules. Autonomous and adaptive schedulers continuously monitor system feedback, learn from past decisions, and proactively respond to failures or performance degradation, reducing manual interactions [3]. But even now, several challenges do exist in terms of energy efficiency and latency reduction. There is still room for enhancing real-time responsiveness for many critical applications, such as healthcare systems [4].

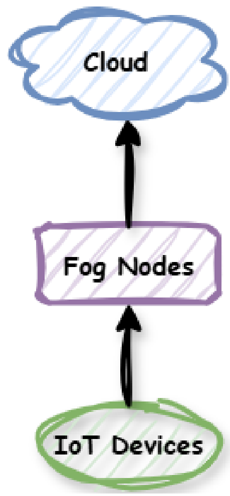

Figure 1. Basic Architecture of IoT-Fog-Cloud Paradigm.

Related work

Li *et. al.* [5] investigated a computing resource scheduling problem in edge-assisted autonomous driving systems where vehicles periodically offload sensor data (localization, obstacle detection, tracking) to a nearby edge server. The authors define a new metric called Age-of-Result (AoR), representing the distance travelled since the last useful computation update for recording processing delays and vehicle mobility. To achieve this, the paper models the scheduling problem as a Restless Multi-Armed Bandit (RMAB) and derives a Whittle index that quantifies the priority of scheduling each vehicle's task. The authors evaluate both synchronous and asynchronous sensing/offloading patterns, and simulation results (including on a real Didi driving dataset) show significant reductions in AoR compared to baseline policies like Highest-AoR-first and Round-Robin. However, the work optimizes only Age-of-Result (AoR) without considering other crucial QoS dimensions such as latency, energy consumption, resource utilization, computation cost, *etc*. The approach further introduces computational overhead, requiring long prediction windows and heavy neural-network training, which is impractical for lightweight, resource-limited Fog nodes.

Lin *et. al.* [6] introduced a metaheuristic-based framework called SPP-DEA to address the Service Placement Problem (SPP) in Fog environments, which is a NP-hard problem. To manage this degree of complexity, the authors modified the Differential Evolution Algorithm (DEA) by incorporating a shared parallel architecture and integrating it into the MADE-k autonomous planning model. While this approach effectively applies DEA for microservice placement, the approach lacks continuous adaptability as the decisions regarding placement are done once per time period, assuming stable task arrival rate and fixed deadlines. Additionally, the article does not incorporate predictive intelligence, limiting the use cases in real-world IoT-Fog-Cloud scenarios.

Table 1. Related Literature Review.

Reference & Year	Approach	Nature of work	Limitations
Li *et. al.*, 2021 [5]	RMAB, Whittle Index, DQN, LSTM, SSA	Autonomous and Adaptive Scheduling	• Homogeneous type tasks are considered • Does not consider delay modeling
Lin *et. al.*, 2023 [6]	Parallel DE (SPP-DEA)	Task Optimization/Resource Placement	• Lacks runtime adaptation • Decision making is centralized.
Choppara *et. al.*, 2025 [7]	DDPG, Actor–Critic	Adaptive Task Scheduling/Resource Optimization	• High computational and training complexity. • Does not consider real-world dataset for validation. • Does not consider cost of task migration.
Choppara *et. al.*, 2025 [8]	FedDQN, K-Means	Autonomous & Decentralized Scheduling	• High communication overhead. • Does not consider real-world dataset for validation.
Nagabushnam *et. al.*, 2025 [9]	Multi-Agent PPO	Adaptive & Multi-Agent Optimization	• Does not consider real-world dataset for validation. • Does not consider cost of task migration.
Jin *et. al.*, 2025 [10]	3C Co-Design Framework	Autonomous & Resource-Oriented Architecture	• Does not consider scheduling algorithm and empirical validation. • Lacks in effective autonomous scheduling approach.

Choppara and Mangalampalli [7] proposed a Resource Adaptive Automated Task Scheduling framework for Fog computing using the Deep Deterministic Policy Gradient (DDPG) reinforcement learning algorithm. Fog nodes operate under limited and fluctuating resources, making static or heuristic scheduling ineffective. To address this, the framework dynamically schedules Fog tasks by learning optimal decisions through continuous state–action interactions. Although the paper presents a strong DDPG-based resource-adaptive scheduling framework, the framework assumes complete visibility of system states, including exact resource availability, dependency graphs, and accurate task parameters, which is often unrealistic in real Fog deployments. While the authors incorporate three scheduling types, the framework does not include inter-node coordination.

Choppara and Mangalampalli [8] proposed a comprehensive vendor-selection framework based on Multi-Criteria Decision-Making (MCDM) techniques to rank and select optimal suppliers in supply-chain environments. The paper integrates AHP (Analytical Hierarchy Process) for determining criterion weights and TOPSIS for ranking vendors based on their closeness to an ideal alternative, ensuring both subjective expert input and objective distance-based evaluation are considered. However, the framework is largely static, assuming that vendor attributes (delivery time, reliability, quality, cost) remain constant over time, while in practice these parameters fluctuate dynamically due to multiple reasons. The model also depends heavily on manual expert assessments for AHP weighting, making it subjective and sensitive to human bias. Additionally, the method does not incorporate real-time monitoring, adaptive re-evaluation, or automated feedback loops, meaning vendor rankings cannot update autonomously when performance changes.

Nagabushnam *et. al.* [9] proposed a Fog-Adaptive Multi-Agent Scheduling Optimization framework called FAMASO for dynamic, heterogeneous Fog–Cloud environments. The proposed approach combines Earliest Deadline First (EDF) with multi-agent Reinforcement Learning (RL), using PPO-based recurrent neural networks (PPO-RNN) to model temporal patterns in task arrivals and adapt scheduling decisions in real time. However, the framework suffers from high computational complexity.

Jin *et. al.* [10] presented Fog-Cloud automation as a transformative reference architecture for building fully autonomous Industrial Cyber-Physical Systems (ICPS). The authors suggested a 3C co-design framework that combines communication, computing and control simultaneously while introducing a Fog-Cloud automation prototype. However, the work remains largely architectural and conceptual, without presenting concrete algorithms, quantitative evaluations, or empirical benchmarks validating the proposed model.

Through autonomous and adaptive scheduling, the above limitations can be addressed directly by considering multi-objective, context-aware, and self-optimizing resource management across the IoT-Fog-Cloud environment. Rather optimizing a single objective, these schedulers simultaneously consider latency, energy efficiency, cost, resource utilization, *etc.*, that proves to be more practical in nature. Moreover, they incorporate energy-aware and cost-aware policies, reducing unnecessary communication, energy consumption, thus improving efficiency and robustness. Table 1 presents a review of the related earlier work carried out in this field.

Research Questions

The research questions focus on essential choices in terms of techniques used, objective function defined, integration of Fog-Cloud environment and autonomous scheduling.

1. ***RQ1*****:** What techniques can ensure adaptability and robustness of the scheduling framework under uncertain or adversarial conditions?
2. ***RQ2*****:** How can Deep Reinforcement Learning (DRL) be leveraged to enable autonomous and real-time scheduling in highly dynamic Fog environments?
3. ***RQ3*****:** How can multi-objective optimization (latency, energy, cost, reliability) be integrated into the DRL-based scheduling framework?
4. ***RQ4*****:** How does the integration of Cloud assistance enhance the overall latency–energy trade-off and scalability autonomous systems under heterogeneous network conditions?
5. ***RQ5*****:** How can the proposed autonomous scheduling framework be applied and validated in real-world use cases (*e.g.*, smart cities, healthcare IoT, and autonomous vehicles)?

Key Contribution

The main contributions of this work can be summarized as follows:

1. An extensive review of earlier literature works (2021–2025) is carried out focusing on adaptive and autonomous scheduling Fog-Cloud paradigm, systematically classifying the review work in terms of their contributions, approach, nature of work, their strengths and limitations.
2. The existing limitations have been highlighted that include homogeneous nature of tasks considered, lack in real-time decision making, adaptation to real-time workload, *etc*.
3. Table 1 presents a structured comparative table that shows various approaches which makes it clear that further research can address the highlighted limitations.
4. Formulated some research questions that focuses on gathering specific information about techniques used, objective function defined, integration of Fog-Cloud environment and autonomous scheduling.

Conclusions

This article presents a systematic review of autonomous and adaptive scheduling approaches for the Fog-Cloud paradigm focusing on latency, energy consumption, resource utilization and cost. Through this article we can conclude the following points while this also paves a way for various future research which can be carried out:

1. Due to the demanding nature of the modern workload, the centralized nature of the traditional Cloud approach is not suitable.
2. Through this survey, it can be considered that autonomous and adaptive scheduling can be implemented to address the limitations that the earlier approaches face in terms of latency, energy consumption, resource utilization and cost.
3. This review also clearly shows that most of the earlier research was carried out using synthetic datasets rather than any real-world dataset.
4. Further research can be carried out by focusing on lightweight adaptive and autonomous schedulers which can be effective in reducing latency, energy consumption and cost while maximizing the resource utilization.

References

1. M. Amadeo, C. Campolo, J. Quevedo, D. Corujo, A. Molinaro, A. Iera, R. L. Aguiar and A. V. Vasilakos, “Information-centric networking for the internet of things: challenges and opportunities”, IEEE Network, vol. 30, no. 2, pp. 92–100, 2016. 10.1109/MNET.2016.7437030.
2. M. Kierzynka, L. Kosmann, M. vor dem Berge, S. Krupop, J. Hagemeyer, R. Griessl, M. Peykanu and A. Oleksiak, “Energy efficiency of sequence alignment tools—software and hardware perspectives”, Future Generation Computer Systems, vol. 67, pp. 455–465, 2017. https://doi.org/10.1016/j.future.2016.05.006.
3. R. Mahmud, K. Ramamohanarao and R. Buyya, “Application management in Fog computing environments: A taxonomy, review and future directions”, ACM Computing Surveys (CSUR), vol. 53, no. 4, pp. 1-43, 2020. https://doi.org/10.1145/3403955.
4. K. H. Abdulkareem, M. A. Mohammed, S. S. Gunasekaran, M. N. Al-Mhiqani, A. A. Mutlag, S. A. Mostafa, N. S. Ali and D. A. Ibrahim, “A review of Fog computing and machine learning: Concepts, applications, challenges, and open issues”, IEEE Access, vol. 7, pp. 153 123–153 140, 2019. 10.1109/ACCESS.2019.2947542.

5. M. Li, J. Gao, L. Zhao and X. Shen, "Adaptive computing scheduling for edge-assisted autonomous driving", IEEE Transactions on Vehicular Technology, vol. 70, no. 6, pp. 5318-5331, 2021. 10.1109/TVT.2021.3062653.
6. Z. Lin, L. Lu, J. Shuai, H. Zhao, and A. Shahidinejad, "An efficient and autonomous planning scheme for deploying IoT services in Fog computing: A metaheuristic-based approach", IEEE Transactions on Computational Social Systems, vol. 11, no. 1, pp. 1415-1429, 2023. 10.1109/TCSS.2023.3254922.
7. P. Choppara and S. S. Mangalampalli, "Resource Adaptive Automated Task Scheduling Using Deep Deterministic Policy Gradient in Fog Computing", IEEE Access, vol. 13, pp. 25969 – 25994, 2025. 10.1109/ACCESS.2025.3539606.
8. P. Choppara and S. S. Mangalampalli, "Adaptive Task Scheduling in Fog Computing Using Federated DQN and K-Means Clustering", IEEE Access, vol. 13, pp. 75466 – 75492, 2025. 10.1109/ACCESS.2025.3563487.
9. G. Nagabushnam, K. H. Kim and Y. Choi, "FAMASO: Fog-adaptive multi-agent scheduling optimization", Cluster Computing, vol. 28, no. 8, pp. 533, 2025. https://doi.org/10.1007/s10586-025-05588-3.
10. J. Jin, Z. Pang, J. Kua, Q. Zhu, K. H. Johansson, N. Marchenko and D. Cavalcanti, "Cloud-Fog automation: The new paradigm towards autonomous industrial cyber-physical systems", IEEE Journal on Selected Areas in Communications, vol. 43, no. 9, pp. 2917 – 2937, 2025. 10.1109/JSAC.2025.3574587.