Mean makespan task scheduling approach for the edge computing environment

Nisha Saini, Jitender Kumar
Department of Computer Science and Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Murthal (Sonipat), India

ABSTRACT
Task scheduling in the edge computing environment poses significant challenges due to its inherent NP-hard nature. Several researchers concentrated on minimizing simple makespan, disregarding the reduction of the mean time to complete all tasks, resulting in uneven distributions of mean completion times. To address this issue, this study proposes a novel mean makespan task scheduling strategy (MMTSS) to minimize simple and mean makespan. MMTSS optimizes the utilization of virtual machine capacity and uses the mean makespan optimization to minimize the processing time of tasks. In addition, it reduces imbalance by evenly distributing tasks among virtual machines, which makes it easier to schedule batches subsequently. Using genetic algorithm optimization, MMTSS effectively lowers processing time and mean makespan, offering a viable approach for effective task scheduling in the edge computing environment. The simulation results, obtained using cloudlets ranging from 500 to 2000, explicitly demonstrate the improved performance of our approach in terms of both simple and mean makespan metrics.

Keywords: Edge computing, Makespan minimization, Processing time optimization, Task allocation, Task scheduling, Virtual machines

This is an open access article under the CC BY-SA license.

Corresponding Author:
Nisha Saini
Department of Computer Science and Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Murthal-131039, Sonipat, India
Email: saini.nisha0203@gmail.com

1. INTRODUCTION
The extensive deployment of internet of things (IoT) applications has led to the development of edge computing, offering low latency and fast response times for time-sensitive applications, including healthcare, emergency services, and traffic monitoring. The edge layer integrates edge and cloud resources to offer services and streamline data flow management [1]. The diversity of edge computing necessitates the use of efficient techniques to effectively optimize user requests or workloads. Load balancing is crucial for improving resource utilization [3], minimizing makespan [4], and optimizing the overall performance of edge computing systems [5]. Two approaches for evenly distributing cloud load monitoring are virtual machines (VMs) and task scheduling. In edge computing, task scheduling is considered an NP-hard problem [6] due to the diverse configurations of hosts and VMs, which can quickly adapt to fluctuating user requests. It is very challenging to identify every prospective mapping among tasks and resources in the edge computing paradigm. Consequently, there is an imperative requirement for an efficient task-scheduling technique that prioritizes the strategic allocation of tasks. The goal is to prevent any VM from being underloaded or overloaded, ensuring an evenly distributed workload across all VMs.

Journal homepage: http://ijece.iaescore.com
Existing heuristics and machine learning approaches designed for task scheduling in edge computing have issues with limited relevance and quick adaptability, which hinder their capacity to solve problems optimally [7]–[10]. Furthermore, these techniques have challenges in identifying temporal workload patterns and are predominantly suitable for centralized deployments. To address these limitations, academics and researchers have diligently investigated several methods to find the optimal solutions in task scheduling [11]–[14]. Shukri et al. [15] developed the enhanced multi-verse optimizer (EMVO) to enhance task-scheduling efficiency. They evaluated the impact of EMVO on makespan time, throughput, and resource utilization. EMVO achieved a reduced makespan and improved throughput and resource utilization compared to the original multi-verse optimizer (MVO) and particle swarm optimization (PSO). However, the performance of EMVO degraded the system performance due to an increase in makespan time. Turkeš et al. [16] performed a meta-analysis on the adaptive layer in adaptive large neighborhood search (ALNS) and identified that it could enhance performance by 0.14%. The meta-analysis studies differ in several characteristics that may impact the significance of the adaptive layer.

Sun et al. [17] emphasized reducing the makespan in workflow scheduling within complex networks in edge computing. They introduced the improved greedy search (IGS) and improved composite heuristic (ICH) methods, surpassing current scheduling algorithms. This technique had a restricted energy capacity for each device and was interdependent among processing tasks. Li et al. [18] introduced a mixed-integer linear programming (MILP) model for task scheduling in a toy-edge-cloud architecture, resulting in notable reductions in solution time. Zhang [19] focused on the concurrent connection between tasks and applied the ant colony method to improve task parallelism and decrease task delays. Although this approach revealed significant global optimization performance, it had a poor convergence pace. Agarwal et al. [20] presented a cloud-distributed scheduling model that considers multiple optimization objectives such as makespan, cost, energy consumption, resource utilization, and load. They designed an expert advisor (EA) using the sine function to achieve the best solutions for this framework. However, despite these algorithms offering benefits in accomplishing optimal solutions with reduced makespan and energy consumption compared to existing approaches, they have limitations in lacking a thorough quantitative examination of the benefits of makespan minimization compared to conventional techniques.

To address these limitations, researchers began to explore the combination and improvement of metaheuristic algorithms. For instance, Vispute and Vashisht [21] devised a hybrid algorithm combining genetic algorithm (GA) and particle swarm optimization (PSO) to minimize energy consumption in user-terminal devices. Mangalampalli et al. [22] presented a heuristic algorithm called GAWOA that utilized genetic and whale optimization algorithms to reduce the total cost. In the context of comprehensive resource scheduling, An et al. [23] introduced an improved reference vector-guided evolutionary algorithm that employed a standard distribution angle penalty distance strategy to select the optimal solution, considering factors such as makespan, cost, load, user expectation, and task completion rate. Wang et al. [24] introduced a two-stage joint optimization model that deals with the flexible job shop scheduling issue and preventive maintenance. This approach helped to save costs in dynamic production environments with frequent product changes. The proposed model has shown effectiveness, but its scalability and broader applicability in other industrial situations need further investigation. The multi-objective crow search algorithm (CSAMOMC) devised by Akraminejad et al. [25] efficiently reduced costs and makespan in scientific cloud operations. It outperforms heterogeneous earliest finish time (HEFT) and time-cost compromise (TC3pop) by an average of 4.42% and 4.77%, respectively. Despite the favorable results, the study predominantly relies on a comparative examination. It fails to meticulously investigate the wider environmental ramifications or contemplate compromises in other facets, such as energy efficiency.

The literature evaluation emphasized the significance of conducting in-depth quantitative evaluations to determine the effectiveness of novel task-scheduling optimization algorithms in reducing makespan compared to traditional methods. The existing literature primarily focuses on minimizing simple makespan through various strategies and algorithms, without addressing the current resource utilization and workload of edge servers. This omission may result in uneven distribution of workloads and inefficient resource utilization, primarily due to the dynamic nature of the computing resources in the edge environments. Based on previous research, we examined and devised a methodology to reduce the simple makespan and processing time by minimizing the mean makespan of individual tasks across all VMS. We designed our approach to establish a task scheduling strategy using the GA, which optimizes parameters such as simple makespan, processing time, optimal resource utilization, and imbalance across VMs by minimizing the mean makespan of individual tasks across all VMs. We have organized this paper as follows: Section 2 explains the proposed research method, while section 3 discusses the findings of this study. Section 4 concludes the paper. To provide readers with a comprehensive grasp of the organization of the work, Table 1. presents a compilation of acronyms that will frequently appear throughout this study.
2. PROPOSED METHOD

The proposed work addresses the challenge of minimizing makespan in edge computing. The main objective of this study is to maximize the efficient utilization of VM capacity in the edge environment. The primary intent is to decrease the simple makespan by strategically allocating tasks among several VMs. The proposed work involves optimizing resource utilization and maintaining a uniform workload distribution. Subsection 2.1 describes the problem formulation for achieving the objective of this study. Subsequently, Subsection 2.2 outlines the proposed strategy.

2.1. Problem formulation

The primary goal of this study is to reduce the simple makespan by efficiently utilizing the VM capacity. This optimization ensures efficient VM utilization and offers additional benefits such as reduced processing times, simple makespan, low imbalance on each VM, simplified batch task scheduling in subsequent iterations, and improved accuracy compared to directly reducing makespan. In edge computing, task scheduling involves allocating tasks to the most suitable VMs in data centres to meet user requirements. Consider a set of n tasks \( T = \{T_1, T_2, T_3, \ldots, T_n\} \), and a set of m virtual machines \( VM = \{M_1, M_2, M_3, \ldots, M_m\} \). Let \( P_{ij} \) be the processing time of task \( i \) on the virtual machine \( M_j \) and \( C_{ij} \) be the completion time of task \( i \) when assigned to the virtual machine \( M_j \). A binary decision variable \( x_{ij} \) representing whether task \( T_i \) is assigned to the virtual machine \( M_j \), with a value of 1 indicating assignment and 0 indicating no assignment. Each task is assigned to each VM, aiming to minimize the fitness of the objective function. Consequently, the proposed strategy minimizes the simple makespan by reducing the mean makespan across all VMs. The formulation of the objective and associated constraints is defined as (1).

\[
\text{minimize } \frac{1}{m} \sum_{j=1}^{m} C_{\text{max}}^{(j)} \quad (1)
\]

subject to \( C_{ij} = \frac{1}{m} \sum_{j=1}^{m} P_{ij} \quad (1a) \)

\( C_{\text{max}} = \max_{1 \leq j \leq m} \{ C_{ij} \mid 1 \leq j \leq m \} \quad (1b) \)

\( \sum_{j=1}^{m} x_{ij} = 1, \forall i \in \{1, 2, 3, \ldots, n\} \quad (1c) \)

\( x_{ij} \in \{0, 1\}, \forall i \in \{1, 2, 3, \ldots, n\}, \forall j \in \{1, 2, 3, \ldots, m\} \quad (1d) \)

Constraints (1a) indicate that each task is assigned to precisely one VM. Constraint (1b) calculates the simple makespan as the maximum completion time across all tasks and VMs. Constraint (1c) ensures that each task is assigned to precisely one VM, and constraint (1d) defines that the decision variable \( x_{ij} \) are binary.

2.2. Mean makespan task scheduling strategy

This section outlines a strategic approach for minimizing the simple makespan in an edge computing environment. Our strategy reduces the mean makespan, which assists in addressing issues related to the
uneven distribution of the mean completion time of tasks. Every random variable approaches the mean and provides a better path for optimal solutions. Our technique stipulates the most effectively utilized elucidation for this problem.

Consider an assortment of $n$ tasks and $m$ virtual machines, where each task is evenly distributed across the VMs. The completion time $C_{ij}$ of the task $T_i$ on each virtual machine $M_j$ can be represented as (2).

$$C_{ij} = \frac{1}{m} \sum_{j=1}^{m} P_{ij}$$  \hspace{1cm} (2)$$

where $P_{ij}$ is the processing time of the task $T_i$ on virtual machine $M_j$. Makespan $C_{\text{max}}$ is the maximum completion time for all tasks. It can be expressed as (3).

$$C_{\text{max}} = \max_{1 \leq i \leq n} \{ C_{ij} \mid 1 \leq j \leq m \}$$  \hspace{1cm} (3)$$

The mean makespan $\bar{C}_j$ is the mean completion time for all tasks. It can be expressed as (4).

$$\bar{C}_j = \frac{1}{n} \sum_{i=1}^{n} C_{ij}$$  \hspace{1cm} (4)$$

On substituting the value of completion time into the mean makespan expression, we get (5) to (7).

$$\bar{C}_j = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{1}{m} \sum_{j=1}^{m} P_{ij} \right)$$  \hspace{1cm} (5)$$

$$\bar{C}_j = \frac{1}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} P_{ij}$$  \hspace{1cm} (6)$$

$$\bar{C}_j = \frac{1}{mn} \sum_{j=1}^{m} \sum_{i=1}^{n} P_{ij}$$  \hspace{1cm} (7)$$

Since each task is evenly allocated to each VM, the values of processing time $P_{ij}$ remain constant $\forall i$ and $j$. Therefore, this expression can be rewritten as (8) and (9).

$$\bar{C}_j = \frac{1}{mn} (mn) (P_{ij})$$  \hspace{1cm} (8)$$

$$\bar{C}_j = P_{ij}$$  \hspace{1cm} (9)$$

This demonstrates that the mean makespan is directly proportional to the processing time of tasks on the VM. Thus, by lowering the mean makespan, the processing time and completion time for all tasks on all VMs will also be minimized. As the simple makespan $C_{\text{max}}$ represents the maximum completion time across all VMs, reducing the completion time $\bar{C}_j$ of any task on any VM will correspondingly decrease the $C_{\text{max}}$. Therefore, achieving a minimum in the mean makespan assures a minimum in the simple makespan.

### 3. RESULTS AND DISCUSSION

In this section, we assess and contrast the performance of the proposed MMTSS strategy in terms of i) simple makespan and ii) mean makespan. We carried out the simulation using CloudSim [26]. We evaluated the proposed strategy using the NASA Ames iPCS/860 log dataset, which contains real-world workloads from [27] that CloudSim recognized. Parameters are discussed in Table 2. The results of our study demonstrate that our approach effectively minimizes the simple makespan and enhances the performance of individual tasks by reducing processing time. Furthermore, our strategy of minimizing the mean makespan leads to a decrease in the overall simple makespan. However, with an increase in the number of tasks, there is a higher degree of volatility. Table 3 represents a comprehensive analysis of the simple makespan (AET) and mean makespan (AET) metrics, which span a range of cloudlet quantities from 500 to 2000. This analysis aims to clarify the distinctions in the mean execution time observed across the two metrics over different numbers of cloudlets. This comparison is optimized by Figure 1, which visually depicts the experimental analysis of simple and mean makespan. It provides a visual depiction of their performance in terms of task allocation on VMs in the edge computing environment.
During our investigation, we observed an unforeseen pattern in our analysis of the correlation between the mean makespan and the measurements of the simple makespan. Generally, we expect the mean makespan values to decrease in comparison to the simple makespan, signifying an overall optimization in the scheduling process. However, the results of our study demonstrated a deviation from the predicted trend, indicating a complete reversal in the expected direction. After carefully analyzing the data, it became clear that the mean makespan readings demonstrated more significant variations than the simple makespan values. This disparity prompted an in-depth investigation of the fundamental elements that contributed to this observed pattern.

The subsequent study revealed a potential relationship between the significant variation in the mean makespan observations and the impact of outlier observations within the dataset. Outliers, which are data points that deviate significantly from many observations, have the potential to disproportionately affect the estimation of the mean. Due to their exceptional characteristics, outliers significantly impact the mean value of a dataset, which in turn increases the observed variability in the mean makespan values. Therefore, the difference between the expected and actual results highlights the significance of carefully examining the data and considering possible anomalies in statistical analysis. Let’s consider a hypothetical scenario where we introduce four tasks, T1 (70 s), T2 (10 s), T3 (5 s), and T4 (15 s). These assignments have a total duration of 100 seconds, resulting in an average or mean value of 25 seconds. There are significant fluctuations, especially at extreme values like 70 seconds. Outliers have an impact on the mean, so devising a solution to address this issue is critical. Hence, developing a system that can effectively reduce the impact of outlier values and provide a more resilient optimization methodology is crucial. Despite these results, further studies will prioritize exploring alternate parameters and approaches to reduce the impact of outliers on the mean makespan computations.
4. CONCLUSION AND FUTURE SCOPE

The paper presents the MMTSS, a novel approach for minimizing makespan in the edge computing environment. This study deviates from previous literature by optimizing parameters such as simple makespan, processing time, and imbalance across VMs. This is achieved by minimizing the mean makespan of individual tasks across all VMs. Our proposed approach aims to alleviate the challenges associated with the uneven distribution of the mean completion time for tasks. The objective is to minimize tasks’ simple makespan and processing time by implementing a GA technique to reduce the mean makespan across all VMs. The method enhances overall performance by optimizing processing times and lowering the simple makespan through a uniform distribution of tasks among VMs. Moreover, the paper highlights the need to effectively utilize VM capacity to ensure adequate workload allocation, minimize imbalance, and enhance task scheduling accuracy in subsequent iterations. As part of future work, we aim to incorporate additional metrics for better convergence of the metaheuristic algorithms. In addition, we also plan to extend the task scheduling strategy to the hybrid cloud environment.

REFERENCES


BIOGRAPHIES OF AUTHORS

Nisha Saini currently immersed in her Ph.D. in computer science and engineering at Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonipat, Haryana, India. She holds academic distinctions with B.Tech. in information technology and M.Tech. in computer science and engineering, both conferred by the same university. She possessed a three-year tenure of teaching experience prior to her current academic pursuits. Her research mainly focuses on the internet of things (IoT), cloud computing, and edge computing. She can be contacted at email: saini.nisha0203@gmail.com.

Jitender Kumar received the M. Tech. degree in computer science and engineering from the Guru Jambheshwar University of Science and Technology, Hisar, Haryana, India, in 2008. He has completed his Ph.D. at Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonipat, Haryana, India. He is an assistant professor at the Department of Computer Science and Engineering, Deenbandhu Chhotu Ram University of Science and Technology. His research interests include cloud computing and mobile computing. He can be contacted at email: jitenderkumar.cse@dcrustm.org.