Accurate Prediction of Workflow using Dual-Stage Learning to Reduce Task Execution Time

N. Bhalaji

Associate Professor, Department of Information Technology, SSN College of Engineering, Kalavakkam, Tamil Nadu, India

E-mail: bhalajin@ssn.edu.in

Abstract

As the number of cloud data centres continues to expand rapidly, one of the biggest worries is how to keep up with the energy demands of all these new servers without negatively impacting system dependability and availability or raising the price of power for service providers. Workflow task performance prediction for variable input data is crucial to several methods, including scheduling and resource provisioning. However, it is challenging to create such estimations in the cloud. The suggested system's two-stage forecasts and parameters that account for runtime data, allow for very precise predictions. The workflow is smooth, and obviously the task execution time is adequate. This strategy beats the state-of-the-art prediction techniques, as shown by empirical data. It is demonstrated that the models of this form, predicting workflow for a given cloud, can be easily transferred to other clouds with little effort and error.

Keywords: Energy consumption, execution time, prediction, learning algorithm, cloud computing, fog computing, run time

1. Introduction

The boundaries of science have shifted dramatically in the modern period. There have been many advancements in areas of digital technology, including social media and networks, money transfers, sensor data, commercial and financial activities, and person-to-person conversations through digital channels. Due to these advancements, huge datasets (sometimes known as "Big Data") have been produced. Pictures, words, XML, music, social context, videos, etc. are all examples of data. The problem is that the quantity of data that has to be stored, processed, and analysed is rising exponentially faster than our ability to do so using

DOI: https://doi.org/10.36548/jismac.2022.4.002
Received: 21.09.2022, received in revised form: 19.10.2022, accepted: 30.10.2022, published: 11.11.2022

standard databases and methods. This problem has sparked interest in developing smart and efficient platforms and methodologies for processing, storing, and analysing the massive amounts of data that exist in almost every industry. Research and businesses alike encounter the greatest difficulty in trying to accurately predict how long a given activity will take to complete in light of the ever-increasing volume of data that has to be analysed [1,2].

1.1 Dynamic Remote Access to the Cloud

To address these issues and the increased data storage and processing requirements of mobile devices, cloud computing has gone mobile [3 -7]. Through the use of mobile cloud computing, one may offload computationally heavy programmes or processes to a distant cloud server, which provides access to vast computational and storage capabilities. On a mobile device, compute-intensive operations often use more power because they conduct several complicated computations on a relatively limited dataset. In addition, many modern jobs cannot be carried out on mobile devices since they cannot be completed in the allotted amount of time. Implementing them on the cloud is useful since the execution time is much less than the quantity of data transmitted. As a result, new migration algorithms are being developed for doing such activities in the cloud. Task migration may be a useful tool but gauging its success can be challenging. This is because moving tasks introduce new expenses, such as those associated with transferring data across a network.

1.2 Cloud Computing Efficiency

The efficiency and low price of cloud computing are very useful for large-scale networks. SCs, much like cloud computing, provide the dynamic de-provisioning of resources in response to demand. This increases operational flexibility. Scheduling, which encompasses resource, job, workflow, task, and deadline scheduling, is one of several elements that determines performance and reliability in the sensor cloud. Scheduling tasks allows to make the most of the available assets and time, guaranteeing a quality end product that will please the clients. Task scheduling in cloud computing is the process of dividing up work amongst several virtual computers so that it may be completed as rapidly as feasible [8]. This study focuses on sensor cloud environments for agricultural irrigation control systems, with a particular emphasis on cloud-user communication. In this setup, requests from end users are seen as tasks that must be completed in order to have access to the necessary data in the cloud. Users may access a variety of digital tools, thanks to the cloud. In order to do their work, users tap into a variety of online tools. Thanks to the scalability of the cloud, many virtual applications may run concurrently.

The resources of multi-goal computers are shared. The distribution of resources is based on what users need. An effective module is for scheduling checks and reports on the availability of resources. Having a reliable scheduler is crucial for real-time efficiency in cloud computing. Using a task scheduling method, cloud-based jobs may be assigned to virtual machines. It makes advantage of the available means. This improves the efficiency of resource use and the throughput of the system while decreasing the time it takes to get a response to a request.

A server cluster on the Internet sharing its resources is the common conception of the cloud. Both physical (memory, processors on the server, etc.) and virtual (code, databases, etc.) assets are counted. The only reason the internal machine can communicate with the outside world is to issue an HTTP request, for all the processing to be completed by the cloud computing providers [9]. However, the widespread nature of cloud computing's resources has made it difficult to make efficient use of those resources, slowing the industry's fast expansion [10]. This is due to factors like increased system load and requirements for dynamic task scheduling.

1.3 Execution of Virtual Machine

Workflow input data, the kind of Virtual Machine (VM) on which the job is executed, and runtime information dependent on the hardware are all factors that go into characterising cloud-based workflow task executions.

1. A novel totally automated two-stage approach for forecasting task execution durations based on diverse input data from different cloud providers was tested on many real-world workflow applications.
2. The machine learning regression approaches like random forest were used for predicting the duration of workflow tasks.
3. A look at how easily the suggested method of forecasting may be applied to other cloud service providers is given.

1.4 Motivation

Unfortunately, the results of using these approaches to manage the scheduling of several tasks are not optimal. Recent decades have seen cloud computing's rise to dominance as it provides consumers with a plethora of options. The rising number of cloud users is a direct result of the widespread use of cloud computing. With an increasing user base, comes a host of new problems. When working in the cloud, it might be difficult to figure out how to plan activities and assign them to virtual machines. The user's task request must be processed by the most sensitive virtual machine in the cloud. The characteristics of energy use, cost, resource usage, execution time, make-up time, throughput, and reaction time are all decreased by this effective method when time constraints are applied. This research reveals a very effective method for managing applications in terms of user demand and priority (time, energy, cost, and deadline), all while enhancing the QoS level.

2. Literature Survey

At this time, people are mostly interested in analysing large amounts of data. Due to the scarcity of resources, most of the data that is given must be used. Data output is increasing at an unprecedented rate. Hadoop MapReduce is a widely lauded system for swiftly and cheaply processing massive amounts of data. To get the most of a computer system though, how each component contributes to the whole must be understood [11].

Due to the diverse nature of cloud resources, task scheduling in a sensor cloud context, like cloud computing, presents a substantial challenge. There are numerous QoS criteria, hence many different scheduling algorithms have been developed to deal with this problem. It is a common practice to classify scheduling approaches as either heuristic, metaheuristic, or algorithmic. The optimal solution in heuristic approaches is found via the use of prediction [12].

The least amount of time and complication are necessitated for scheduling tasks. Metaheuristic approaches, in contrast to heuristic techniques, pinpoint the most effective solution outcomes. A few of the most cutting-edge methods for scheduling tasks are outlined here. Environments and procedures inside sensor clouds that use minimal amounts of energy were investigated. Background anomaly was detected using computationally cheap algorithm. Most studies neglected to consider how energy efficiency can affect quality of service, scalability, or the lifespan of a network. Quality of service (QoS) and energy efficiency should be maximised in real-time applications such as those used in agriculture, healthcare, and smart homes. Academics may use the findings of this research to improve their methods of measuring QoS while conserving energy [13].

On-field sensor networks, according to Ojha et al., scheduling tasks might benefit from an energy-efficient technique called dynamic duty scheduling. The processing demands of field WSNs are lowered by the sensor cloud architecture. The authors showed a method for selecting appropriate time intervals to upload data to the cloud during off-duty times. The idea suggested saving money through increasing energy efficiency. Traditional energy efficiency, network lifetime, cost-effectiveness, and utility were all surpassed by the suggested technique. Testing is necessary for real-time sensor cloud applications, and one such metric is the QoS assurance [14].

To lower energy usage while connecting sensor networks to sensor clouds, Chatterjee et al., refined the selection process for picking appropriate bridge nodes. Multi-hop data transfer from PSNs to sensor clouds is a current area of study. Saving energy was one option. Taking into account node heterogeneity, mobility, and other network variables is important for testing real-time applications [15].

To meet the QoS requirements of the end-user, Dinh et al., designed a model of interaction between sensors in the cloud. In order to meet the needs of QoS between actual WSNs and cloud-based sensors, a feedback control system was developed. The sensor's energy efficiency was improved via feedback. When compared to more conventional techniques, this method significantly reduced delay. It was said that it is important to test a real-time application. Low signalling overhead is an area of study related to improving energy efficiency [16].

Proshikshya Mukherjee argued that task scheduling is particularly important in cloud and sensor-based systems. This author investigated the problems and obstacles of sensor cloud job scheduling. Time management and organisational skills are essential for success in this position. Researchers will learn a great deal about how to schedule tasks as a consequence of the study. An effective hybrid task scheduling method must be developed for this purpose. For optimal results in the sensor cloud, it is important to make decisions based on a number of factors at once [17].

Swagatika et al., conducted an in-depth analysis of popular scheduling techniques for effectively allocating VMs in the cloud. To foretell how resources would be used, a modified Markov chain model was used. The cloud used a more advanced PSO algorithm to allocate its resources optimally, with dynamic load balancing based on the VM allocation method.

Consideration of several factors, including makespan time, energy use, and cost, was required [18].

2.1 Context of the Problem

Job scheduling in virtual machines is the subject of this study. The primary responsibility of scheduling is to fairly allocate costs by considering how various resources associated with a given task translate to various resources available on a given virtual machine.

3. Methodologies

In this research, a new approach to estimate how long it will take to complete workflow tasks of given different sets of parameters has been presented. These runtimes are represented as functions which take into account both the workflow's inputs and the cloud's capabilities. Such models are developed from past cloud-based iterations of the process using regression techniques. The characteristics of the VM type on which the job is run are what are referred to as "cloud features". Considering that a virtual machine's execution time might vary depending on the cloud in which it was deployed, how long it takes to complete a task after it was started is tracked. In order to determine how closely a given VM and a given task's execution time match, a two-stage prediction process is used. If this work has been performed previously, then may be this information can be retrieved from a database. In such case, the workflow input data and the VM type will be used in a regression model to predict the runtime parameters. The inputs in the second stage's final regression approach forecasts how long it will take to finish the job [19].

3.1 Significance of proposed work

The parameters under consideration are divided into two categories: those that may be set before execution and those that can only be adjusted during execution. Before a job is launched on the cloud, its initialization settings may be statically specified. Executing a job on a variety of virtual machines from the same or various cloud providers allows to accurately gauge how each machine will perform. Some examples of these metrics include user CPU use, system CPU usage, the number of blocks written by a task to memory, and the quantity of data sent by the mission over the network. Figure 1 shows the block diagram of the proposed workflow forecasting or prediction system.

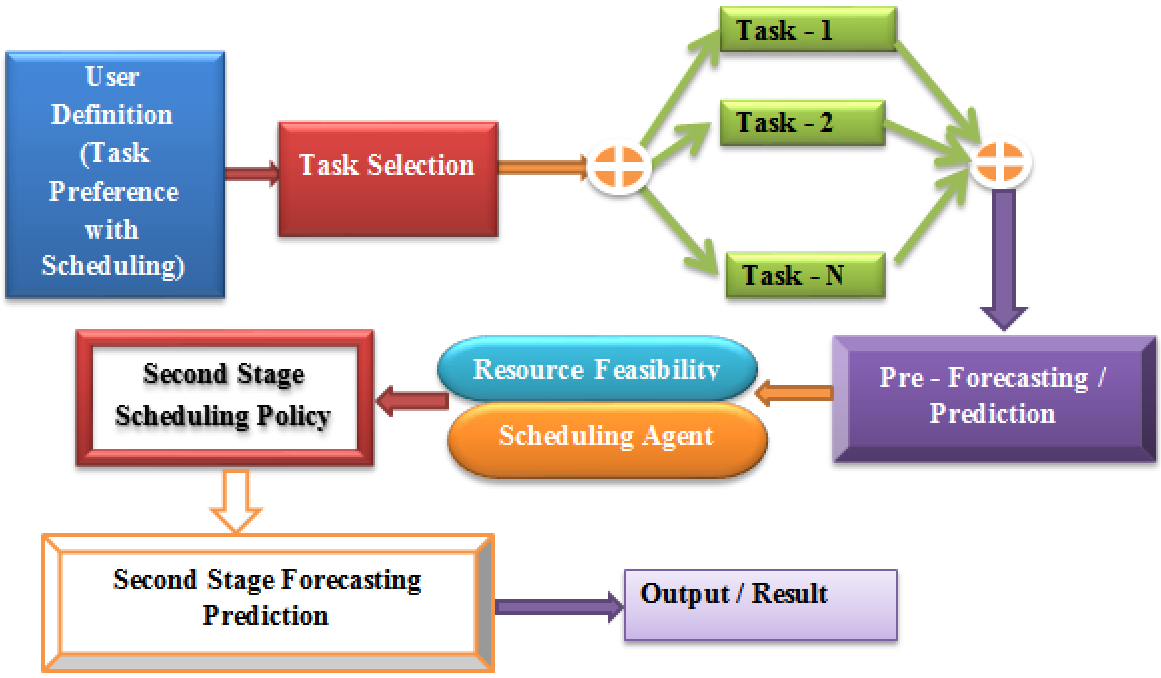

Figure 1. Proposed Workflow of Forecasting/Prediction System

3.2 Workflow

Initially, a model is created for each of the parameters that are used throughout execution. These models are built with Machine Learning (ML) methods in mind, and their goal is to be able to predict with accuracy what hardware a cloud will use to operate a particular VM. An ML technique that connects process input to VM type information and runtime settings gains this understanding indirectly. In the second phase, the models of runtime parameters created in the previous phase are used to make an estimate of how long a job will take to complete [20].

Workflow input is used as training data, much as in similar works. Since changing the workflow input might affect the execution timeframes of the functions making up that workflow, knowing this is necessary for estimating how long a job will take to complete. The cloud service provider uses these settings to allocate hardware resources to each VM instance, which in turn affects how long it takes to complete a given job. Since these values are only accessible via a cloud-based deployment, they are referred as "runtime parameters." The amounts of I/O operations, network bandwidth, CPU user time, CPU system time, and more are all included in this dataset [21]. In order to give reliable time estimates for process activities in the cloud, an innovative offline method is presented.

ISSN: 2582-1369

3.3 Second-stage forecasting

The algorithm represents the preliminary step of the forecasting process. This method takes as input the job, the kind of virtual machine, and the cloud for which a runtime parameter prediction is needed. The algorithm then determines the prediction through machine learning technique educated on historical data with just pre-runtime parameters [22]. In this stage, any state-of-the-art machine learning regression technique can be applied.

4. Comparative Observations

As there are currently few methods for scheduling tasks in sensor clouds, a strong emphasis on the communication between users and the cloud infrastructure has been placed in this study. The makespan time, resource utilisation, and deadline limitations, were taken into account in this investigation. The article has discussed how the increasing quantity of cloud servers is being brought about by the increased processing demands of clients. However, the power consumption of these servers is rather high. Power consumption is a major problem in sensor and cloud environments alike. As a result, optimising energy use through scheduling tasks is crucial for lowering energy consumption and enhancing other factors. Table 1 contains some results obtained through computations.

Table 1. Results obtained

Method	Energy Consumption	Execution Time (sec)	Prediction Rate (%)	Prediction Error
Without Workflow Assignment	Higher Energy	2671.21	61.54%	2.93
Single stage Scheduling	Moderate	814.93	80.39%	0.25
Dual Stage Forecasting	Very Low Consumption	489.12	97.35%	0.004

Based on a rapid examination of the data, it is concluded that Random Forest is the most effective regression technique for reducing prediction errors in both one- and two-stage procedures. The two-stage method achieves inadequate prediction errors in all but one job of the Montage process as compared to single-stage procedures, which is why utilising RF instead is proposed. When comparing the two methods, the single-stage method employing solely pre-runtime parameters performs somewhat better. Good forecasts can be made using only the parameters set up in front for this assignment. However, it is found that other work predictions may be much improved by making use of runtime parameters. Figure 2 shows the graph between energy and execution time.

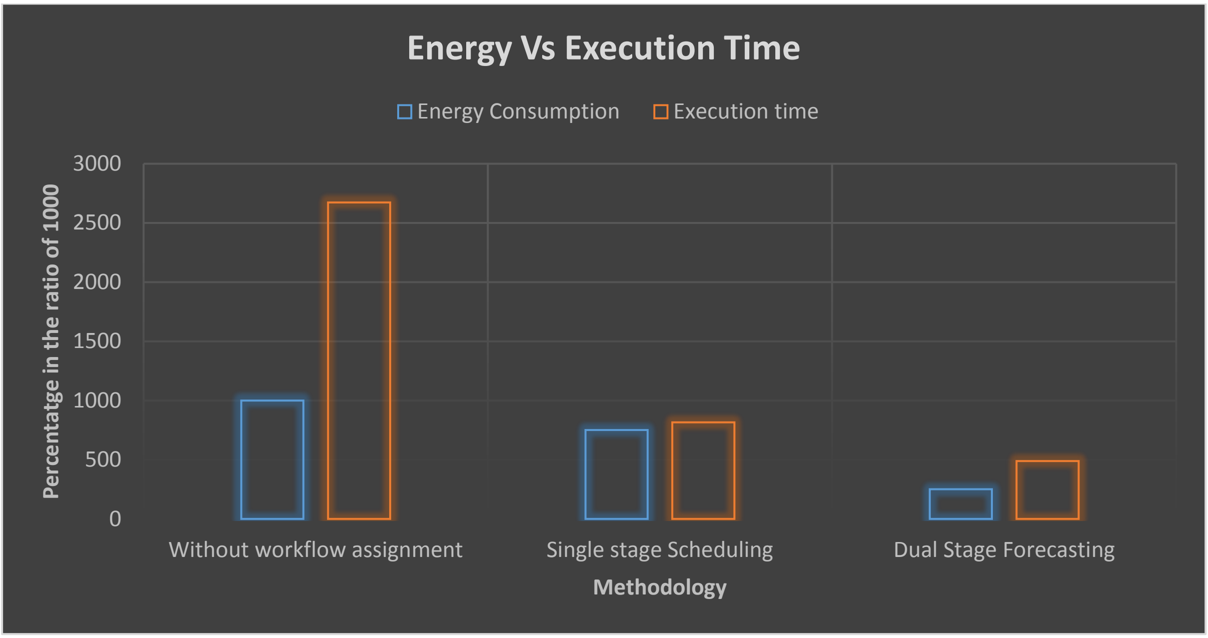

Figure 2. Energy vs Execution Time

When looking at all workflow tasks that have an average reported prediction error for each workflow, the experiment demonstrates that this two-stage strategy is superior to the single-stage approach in all circumstances. Figure 3 shows the comparison between prediction rate vs error.

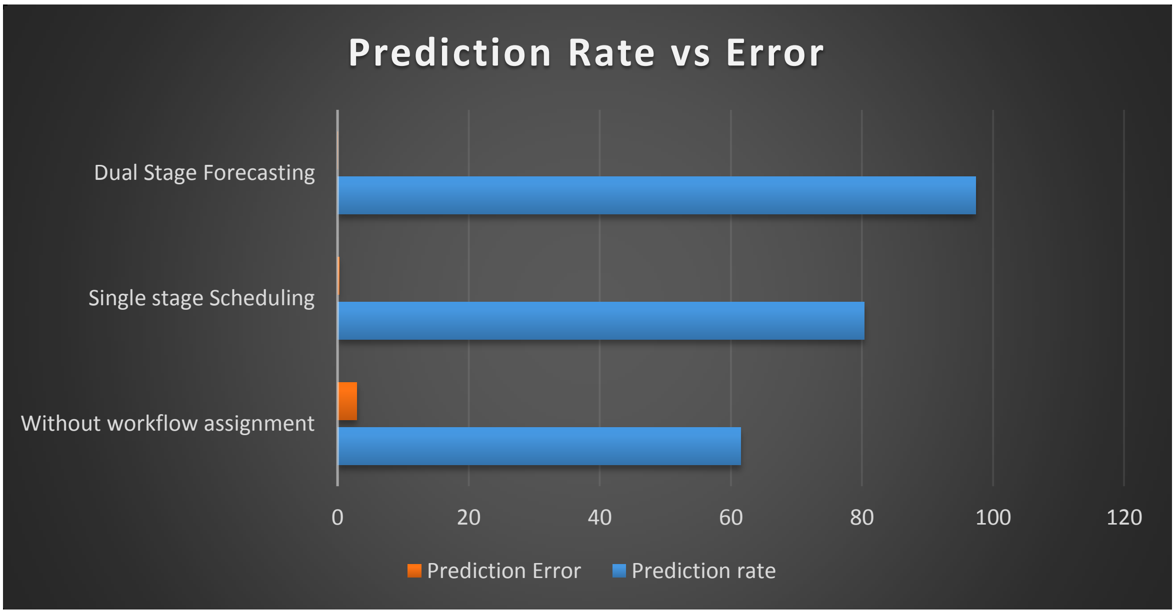

Figure 3. Prediction Rate Vs Error

This holds true regardless of the regression technique used. In the case when the two-stage procedure employs LR as the regression technique, the RAE values computed by the single-stage approach and the proposed approach are identical. This RF technology is improved upon previous methods by reducing RAE by a factor of three when compared to those that just use a single stage. Gains in the Blender process may be as high as five times than the prediction error.

5. Conclusion

This research makes a two-stage prediction of the time to be taken to complete a job on a certain cloud, based on the input data provided. Prior cloud or cloud data for the specific job is used to suggest values for the runtime parameters. The second phase incorporates both the runtime and pre-runtime factors into a time estimate for the task's completion. This work has devised two algorithms that leverage all available fog resources in the system and complete tasks in close proximity to IoT devices to address the aforementioned problem. The results of the experiments show that both the execution time and the power consumption of the cloud servers may be reduced by 20%. However, there has been a 12% rise in the amount of electricity required to run fog infrastructure.

5.1 Possible Future Assignments

As a first step in developing this predictor, it collects a large amount of training data and devises a model for making predictions. For highly variable cloud workloads, this strategy may make sense. In order to solve this issue, this might do an update of predictor when each job is completed. The study will also examine how the proposed model may be used to facilitate various scheduling and resource provisioning strategies.

By performing preprocessing operations closer to the end-user, fog computing technology enables reduced latency, decreased power consumption on the cloud side, and great scalability. However, there are times when the quantity of requests exceeds the network's capacity, or when certain resources in the fog computing network are inappropriate for certain tasks. Consequently, minimising the number of tasks sent to the cloud is the most effective course of action. As a result, additional fog resources are sitting idle; federating them is preferable for sending them to a cloud server. Because of the nature of huge data and time-sensitive applications, this problem has a direct impact on the efficiency of the fog environment. In order to organise tasks generated by Internet of Things (IoT) gadgets, this study seeks to develop a fog topology job scheduler to discover all the fog nodes and their capabilities.

References

[1] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, T. R. Gadekallu, and G. Srivastava, “Sp2f: a secured privacy-preserving framework for smart agricultural unmanned aerial vehicles,” Computer Networks, vol. 187, p. 107819, 2021.

[2] N. Deepa, M. Z. Khan, B. Prabadevi, D. R. Vincent Pm, P. K. R. Maddikunta, and T. R. Gadekallu, “Multiclass model for agriculture development using multivariate statistical method,” IEEE Access, vol. 8, pp. 183749–183758, 2020.

[3] L. Abualigah and A. Diabat, “A novel hybrid antlion optimization algorithm for multi-objective task scheduling problems in cloud computing environments,” Cluster Computing, vol. 24, no. 1, pp. 205–223, 2021.

[4] J. Praveenchandar and A. Tamilarasi, “Retracted article: dynamic resource allocation with optimized task scheduling and improved power management in cloud computing,” Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 3, pp. 4147–4159, 2021.

[5] P. K. Bal, S. K. Mohapatra, T. K. Das, K. Srinivasan, and Y. C. Hu, “A joint resource allocation, security with efficient task scheduling in cloud computing using hybrid machine learning techniques,” Sensors, vol. 22, no. 3, p. 1242, 2022.

[6] M. Gokuldhev, G. Singaravel, and N. R. Ram Mohan, “Multiobjective local pollination-based Gray Wolf optimizer for task scheduling heterogeneous cloud environment,” Journal of Circuits, Systems, and Computers, vol. 29, no. 07, pp. 2050100–2050124, 2020.

[7] K. Dubey, M. Kumar, and S. C. Sharma, “Modifified HEFT algorithm for task scheduling in cloud environment,” Procedia Computer Science, vol. 125, pp. 725–732, 2018.

[8] Y. J. Moon, H. C. Yu, J. M. Gil, and J. B. Lim, “A slave antsbased ant colony optimization algorithm for task scheduling in cloud computing environments,” Human-Centric Computing and Information Sciences, vol. 7, no. 1, p. 28, 2017.

[9] W. Deng, J. Xu, and H. Zhao, “An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem,” IEEE Access, vol. 7, pp. 20281–20292, 2019.

[10] Y. Su, Z. Bai, and D. Xie, "%e optimizing resource allocation and task scheduling based on cloud computing and Ant Colony Optimization Algorithm," Journal of Ambient Intelligence and Humanized Computing, vol. 12, p. 0123456789, 2021.

[11] L. Hamami and B. Nassereddine, "Application of wireless sensor networks in the fifield of irrigation: a review," Computers and Electronics in Agriculture, vol. 179, p. 105782, 2020.

[12] A. Bhoi, R. P. Nayak, S. K. Bhoi et al., "IoT-IIRS: internet of %ings based intelligent-irrigation recommendation system using machine learning approach for efficient water usage," PeerJ Computer Science, vol. 7, pp. 5788–e615, 2021.

[13] M. Zhang and S. Xiong, "Sensor-cloud based precision sprinkler irrigation management system," in CWSN 2019, C. S. Guo, Ed., Springer Nature Singapore Pte Ltd, 2019.

[14] T. Ojha, S. Bera, S. Misra, and N. S. Raghuwanshi, "Dynamic duty scheduling for green sensor-cloud applications," in Proceedings of the International Conference on Cloud Computing Technology and Science, pp. 841–846, CloudCom, 2015-February.

[15] S. Chatterjee, S. Sarkar, and S. Misra, Energy-Efficient Data Transmission in Sensor-Cloud, pp. 1–6, IEEE, China, 2015.

[16] T. Dinh and Y. Kim, "An efficient sensor-cloud interactive model for on-demand latency requirement guarantee," IEEE International Conference on Communications, vol. 6, pp. 1–6, 2017.

[17] T. S. Proshikshya Mukherjee, Issues Of Some Task Scheduling Strategies On Sensor Cloud Environment (Smart Inte), Springer Nature, Singapore, 2020.

[18] S. Swagatika, A. K. Rath, and P. K. Pattnaik, "Markov chain model and PSO technique for dynamic heuristic resource scheduling for system-level optimization of cloud resources," ARPN Journal of Engineering and Applied Sciences, vol. 13, no. 3, pp. 1042–1052, 2018.

[19] J. Jayakumar, S. Chacko, and P. Ajay, "Conceptual implementation of artificial intelligent based E-mobility controller in smart city environment," Wireless Communications and Mobile Computing., vol. 2021, pp. 1–8, 2021.

[20] X. Ren, C. Li, X. Ma et al., "Design of multi-information fusion based intelligent electrical fire detection system for green buildings," Sustainability, vol. 13, no. 6, p. 3405, 2021.

[21] H. Tang, C. Li, J. Bai, J. H. Tang, and Y. Luo, "Dynamic resource allocation strategy for latency-critical and computation- intensive applications in cloud-edge environment," Computer Communications, vol. 134, pp. 70–82, 2019.

[22] B. Zade, N. Mansouri, and M. M. Javidi, “Saea: a securityaware and energy-aware task scheduling strategy by parallel squirrel search algorithm in cloud environment,” Expert Systems with Applications, vol. 176, no. 1, article 114915, 2021.

[23] K. Liu, X. Wang, and Z. Qu, “Train operation strategy optimization based on a double-population genetic particle swarm optimization algorithm,” Energies, vol. 12, no. 13, p. 2518, 2019.

[24] A. A. Hussain, B. A. Dawood, and F. Al-Turjman, “IoT and AI for COVID-19 in scalable smart cities,” in International Summit Smart City 360° , pp. 3–19, Springer, 2021.

Author’s biography

N. Bhalaji has more than 15 years of teaching experience. He received his B.E. & M.E. degree in the discipline of Computer Science and Engineering and Ph.D specializing in Trust Based Routing approach for MANETs from Anna University, Chennai. His current research interests include Application of Trust over information and several communication domains namely Internet of Things and Blockchain Technologies. He is a recognised supervisor of Anna University and also Doctoral Committee member for VIT, SRM and Sathyabama University. He is a member of the board of studies in SRM Valliammai Engineering College and Vels University, Chennai.