A Dual-Objective Approach for Allocation of Virtual Machine with improved Job Scheduling in Cloud Computing

Sandeep Sutar
Department of Computer Science and Engineering, Don Bosco Institute of Technology, India
sutarsandeep07@gmail.com

Manjunathswamy Byranahallieraiah
Department of Computer Science and Engineering, Don Bosco Institute of Technology, India
manjube24@gmail.com

Kumarswamy Shivashankaraiah
Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, India
kumar.aruna@gmail.com

Abstract: *In Cloud Computing (CC) environment, requests of user are maintained via workloads that are allocated to Virtual Machines (VMs) using scheduling techniques which primarily focus on reducing the time for processing by generating efficient schedules of smaller lengths. The efficient processing of requests also needs larger usage of resources that incurs higher overhead in the form of utilization of energy and optimization of cost utilized by Physical Machines (PMs). Assignment of VMs optimally in the environment of CC for jobs submitted by users is a challenge. In order to obtain better solution involving scheduling of jobs to VMs, considering two parameters utilization of energy and cost, we present a dual-objective approach for VM allocation with improved scheduling of jobs in CC environment. The proposed work aimed to build a dual-objective scheduling model for improved job scheduling, focusing on minimization of cost and utilization of energy at a time. For evaluating performance of dual-objective approach, we utilized two types of benchmark datasets and compared with existing approaches such as Whale, Artificial Bee Colony (ABC), Particle Swarm Optimization (PSO) and Metaheuristic Dynamic VM Allocation (MDVMA) techniques. The results obtained from simulation demonstrated that dual-objective approach performs better in the form of minimization of utilization of energy and cost.*

Keywords: *Cloud computing, job scheduling, VM scheduling, cost optimization, energy utilization, dual objectives.*

Received December 24, 2022; accepted May 23, 2023
https://doi.org/10.34028/iajit/21/1/4

1. Introduction

Cloud Computing (CC) keeps on altering the advanced world through its proficient practical IT solutions. It has changed the method of computing in our everyday existence by giving storage and resources for computations to its clients. It started through advancement of equipment, distributed computing, Internet, and management of frameworks. Despite the dependency on conventional and parallel computing models [4], it keeps on assuming its part for the state-of-the-art advancements like edge computing and Internet of Things (IoT) [10]. Involving multi-occupancy for sharing of resources and flexibility for dynamic scaling of resource assignment makes CC achieve a satisfactory use of resources for any organizations possessing the cloud [1]. By and large, the environment of clouds offers services via various flavours such as Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [33].

With higher utilization of resources, corresponding cost will also advance. Also, the expanded data center deployments in CC environments with strong servers has maximized utilization of energy. It has in turn increased cost as well as polluted global environment through emissions of excess carbon contents [12, 21], hence making it deficient for green environment. Various strategies of scheduling for productively handling customer demands by mapping them to Virtual Machine (VM) are existing in a cloud-based environment. The pictorial representation of a typical CC environment along with physical equipment associated into datacenters given in Figure 1, where every Physical Machine (PM) is described with CPUs or cores, RAM, storages, power and bandwidths. The use of datacenters resources results in utilization of energy and costs.

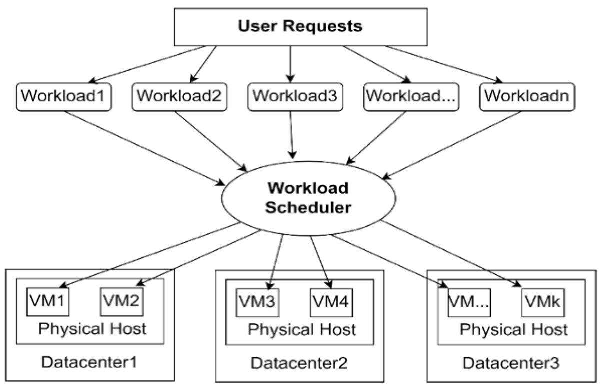

Figure 1. Overview of typical CC environment.

The VMs possessing virtualized resources are allotted to PMs available in datacenters. This assignment is performed by taking in to consideration the expected virtual resources by a VM and resources available in host. The demands of customers are arranged as cloudlets, are executed by VMs. The cloudlets mapping to satisfactory VM is consequently critical for handling requests from users proficiently. Notwithstanding the efficient reaction to requests of users [14], scheduling mechanism should also target limiting costs related to usage of resources and the utilized energy by PMs hosting VMs. A few techniques in [3, 22, 34] likewise aim reduction of cost in cloud environment via compelling strategies of scheduling. Nonetheless, these mechanisms don't focus to simultaneously reducing of energy utilization to help green CC environment.

An effective method for scheduling of job is considered as perhaps the most critical technique to address issues of reduction of utilization of energy and cost in datacenters of the cloud. Job scheduling technique tries to schedule jobs submitted by users to explicit VMs determined to streamline the goals such as maintenance of throughputs, energy utilization minimization while guaranteeing service level agreements and service qualities required [23].

A few techniques have been created to address scheduling of jobs involving heuristic and metaheuristic techniques. Heuristic techniques are utilized to designing of a particular type of scheduling issues, while metaheuristic strategies are utilized for obtaining close to ideal solutions. Metaheuristic strategies could be utilized in three ways to take care of job scheduling issues like single objective scheduling, multi-level objective as single objective, and multi-level objective as multi-objective scheduling techniques [8, 27]. Single objective techniques, at a time, tries to upgrade just a single objective from a bunch of objectives. Multi-level objective techniques try to enhance various objective capacities by consolidating them into a solitary objective process.

In our work, primarily focused to formulate job scheduling issue as multi-objective issue, whereas existing works majorly concentrated on metaheuristic techniques for the optimization of multiple or single objectives during cloud-based scheduling of jobs. We concentrate on building a metaheuristic model based on multiple objective strategy in obtaining optimized solutions for scheduling jobs guarantying minimized cost and utilization of energy for the execution of set of jobs in cloud VMs. Benchmark data sets are used for evaluating our work and also the results are analysed and compared with some of the methodologies existing.

2. Related Work

Numerous researchers worked on heuristic techniques for taking care of problem of scheduling jobs in CC environment. The heuristic techniques basically concentrate on obtaining an ideal or nearly optimized solutions [13]. Those techniques, which are problem-based [9], investigate search space for tracking down the better solutions by making use of problem's features. Many of heuristic techniques have been created and utilized in cloud environment for scheduling the jobs which are not dependent.

Consolidation of workload is a stage in Cloud server where jobs are designated among accessible hosts so that an insignificant number of hosts is utilized and clients require as far as service level arrangement is satisfied. To accomplish consolidation of workloads, hosts are split between three categories in light of their usage specifically normal, underloaded and overloaded hosts. The Patel and Patel [26], presented host utilization aware technique for underloaded host recognition and putting its VMs on different hosts in a unique CC environment and contrasted this technique with existing ones and with experimental investigations. Works in paper [19] also expects to give an image of energy productivity in distributed computing and also groups heuristics-based enhancement strategies and the unique procedures for energy management.

A few researchers have presented different energy-productive techniques for diminishing the energy utilization in datacenters. Nature-inspired technique is one of them. Work in paper [32] presents an extensive survey of best-in-class Nature-inspired techniques recommended for addressing energy problems in cloud datacenters. A scientific categorization is followed concentrating on three critical aspects in the survey containing energy awareness, consolidation and virtualizations.

The work in [35] set forth an energy productive based job scheduling issue subject to general job delay in view of alternating direction method of multipliers, in a network fitted with edge nodes as well the cloud. The work likewise tended to the concerns on protection disclosures in the transmission of information among IoT gadgets and adapt differential security to many-sided optimization related issues. In order to establish a superior solution for load balancing challenges in CC, having broad information, a hybrid framework was proposed [18] that makes categorization on quantity of documents present in cloud utilizing formatting of file types. The classification is achieved utilizing support vector machine approach [5], by taking in to consideration different document formats, for example, sound, video, texts, and pictures in cloud. Targeting on cloud applications which are computation-based, the work at [20] presented insightful strategies to settle down the two open issues.

On comparative lines, multi-objective Whale optimization technique is created for scheduling of job for a CC environment in [28]. Reddy and Kumar [28] tried to scheduling jobs based on parameters of fitness, depending on three circumstances, to be specific, service quality, energy and usage of resources. They showed that by taking in to consideration these three boundaries, execution time of job and cost of VMs could be limited by utilizing multi-objective Whale Optimization technique. The productivity of their proposed strategies relies on parameters of fitness. Sreenu and Sreelatha [30] proposed a methodology called W-scheduler for tackling job scheduling issue utilizing a coordinated methodology of multi-objective framework and whale optimizing technique. W-scheduler includes the calculation of value of fitness by calculating cost-based functionality of memory and CPU.

Sanaj and Prathap [29] presented a structure for map-reduce and a coordinated methodology for proficient scheduling of jobs in light of genetic methods and whale optimization technique. Initially proposed to extricate job related features, minimised further and pool of jobs are partitioned into various sub jobs utilizing map-reduce mechanism. Jena [17] proposed an Artificial Bee Colony (ABC) methodology for improving energy use, cost, processing time and resource utilization in CC environment. Jena [17] suggested to appoint a timestamp to jobs submitted by users on arrival and keep a queue to serve jobs on the basis of first in first out. Jobs are allotted to appropriate datacenters by utilizing multi-objective ABC technique.

The works at [31] suggested a Particle Swarm Optimization (PSO) approach in view of simulation-based social conduct of bird running to take care of multi-objective improvement issue. Suresh *et al*. [31] termed every solution as a particle and changed its location in search page according to its experience of flying and adjoining particles. Omkar *et al*. [24] suggested an incorporated methodology in light of quantum acted PSO and vector assessed PSO techniques for tackling multi-objective optimization issue as single objective issue. They tracked down suitable loads to objective functionalities to calculate fitness value as a solitary objective value. However, this arrangement needs ensuring the convergence globally and that is settled utilizing quantum acted PSO technique that make use condition of particle with wave functionality rather than speed and location of particles. Alsadie [2] present a Metaheuristic Dynamic VM Allocation (MDVMA) technique for job scheduling in a CC environment. MDVMA concentrate on building a metaheuristic technique for improving job scheduling with target of reducing utilization of energy, cost and makespan at a time to boost service providers of clouds according to the needs.

In our proposed work, in this paper, initially we frame mathematical representations for calculation of cost and energy utilization of scheduling a job in datacenters of cloud. The problem of scheduling a job is framed as dual-objective optimization problem containing utilization of energy and cost as objectives with respect to the user constraints such as deadline of run time. In order to reduce cost and utilization of energy at a time, non-dominated sorting genetic technique [7] is utilized to perform dual-objective optimization for scheduling a job. The simulation of work is done in CloudSim simulating environment, making use of benchmark datasets and results are compared with existing four approaches such as Whale, ABC, PSO, and MDVMA that are discussed above.

3. Proposed Method and Datasets

In this section, methodologies to implement the contribution of our work is discussed. The proposed dual-objective framework consists of several components and steps. Firstly, the problem formulation is discussed, including the definition of objectives, identification of constraints and variables. The mathematical model is designed to integrate the objectives, decision variables, and constraints. An optimization algorithm is chosen and customized to solve the dual-objective optimization problem efficiently. The framework is evaluated using benchmark datasets, and a comparative analysis is performed with existing approaches to demonstrate its superiority.

The method proposed is pictorially represented in Figure 2, containing four primary components, involving optimized utilization of energy, expenses incurred for services of the cloud, and makespan, for the allocation of jobs J={Job_1, Job_2, Job_3, … Job_n} upon the virtual machines V={VM_1, VM_2, VM_3, … VM_k} in the environment of CC. The proposed framework contains three major functional components such as calculation of energy utilization, Estimation of Jobs and Jobs scheduling.

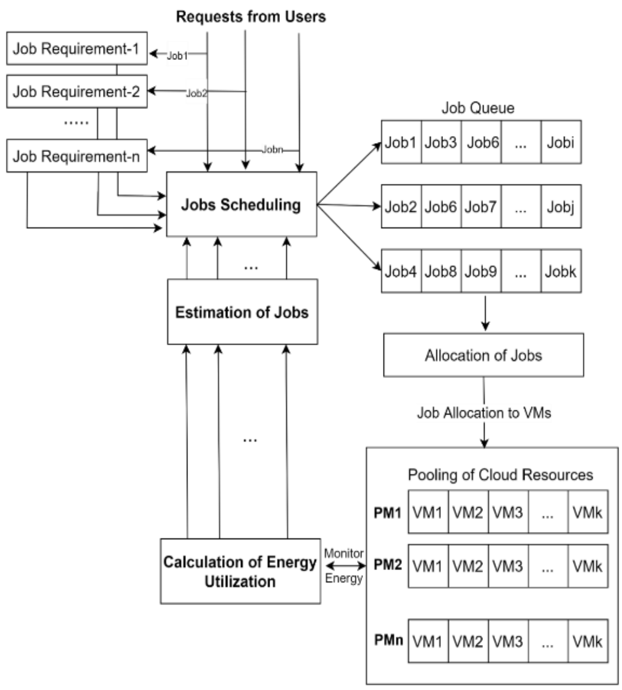

Figure 2. Proposed dual-objective framework.

The calculation of energy utilization component will make use of dynamic voltages frequency scaling-DVFS technique [2] for calculating energy utilization of VMs in the environment of cloud. This technique gives utilization of energy by every resource executing at various frequencies of voltage, described mathematically in Equation (1) below.

$$Energy\ Utilization = \beta \times Voltage^2 \times Freq_{Work} \tag{1}$$

The utilization of energy is calculated utilising voltage level of every piece of jobs Job_i, working frequency and a constant, a factor of frequecny and capacitance of load, with varying values between 0 and 1. The overall utilization of energy for completion of job is calculated using a mathematical model given in Equation (2) below, where $CTime_j$ is time of completion for the job Job_i for a given VM.

$$Overall\ Energy\ Utilization \sum_{j=1}^{k} \beta \times Voltage^2 \times Freq_{Work} \times CTime_j \tag{2}$$

The work done here also concentrate on building a model for scheduling of jobs depending on evaluating parameters such as energy utilization, completion time of job, cost, makespan, waiting time of job and rate of failure [15], for the optimization of multiple objectives. In this work, we intend to concentrate on energy utilization and cost as metrics for evaluating the technique proposed.

The Estimation of jobs component calculates the utilization of energy using mathematical model given in Equation (2).

The cost is amount of spending computational resources for the execution of a job, which is mathematically modelled using Equation (3), with time for execution of the job (TE_j), cost of computational resources per time unit (CR_j), time for completion (TC_j) and time required for transferring the job (TT_j).

$$Cost = \sum_{j=1}^{n} (TE_j \times CR_j) + \sum_{j=1}^{n} (TC_j \times TT_j) \tag{3}$$

One more important component in proposed framework, Jobs scheduling, performs optimal scheduling of jobs J={Job_1, Job_2, Job_3, …, Job_n} upon the virtual machines V={VM_1, VM_2, VM_3, …, VM_k}, and optimizes objectives by establishing the interactions between Jobs and VMs. This component targets to schedule jobs to VMs by achieving lower utilization of energy and costs based on resource availabilities and capacity of PMs. Along with user job requests, the Jobs scheduling component receives as input, utilization of energy and costs, and appropriate job schedules are obtained on particular VMs by applying metaheuristic mechanism.

As a part of optimization technique, obtaining better solution, require evaluating criteria to analyse entire solutions feasibly in searching space. The environment of CC may contain multiple criteria for evaluation that requires to be simultaneously optimized, such as reduction of energy utilization and reduction of cost. This job scheduling issue is formulated as a dual-criteria optimization problem which makes optimization of energy utilization and cost synchronously. An advanced form of non-dominated sorting genetic mechanism [2] is applied for obtaining optimized solutions for dual-criteria job scheduling problem. The dual-criteria functions can be mathematically modelled as given below.

$$Reduce\ (Utilization\ of\ Energy) = \sum UE_{jk} \tag{4}$$

where UE_{jk} is utilization of energy for running job Jj on VM V_k.

$$Reduce\ (Cost) = \sum_{j=1}^{n} (TE_j \times CR_j) + \sum_{j=1}^{n} (TC_j \times TT_j) \tag{5}$$

As per the required constraints of system design, every job *J* must be allocated to one VM, Job's deadline must be fulfilled by time of completion of every job, VM's maximum capacity should not be exceeded by the resource need of entire jobs on a VM. The Allocation of Jobs component in proposed framework, allocates the schedule selected to VM, with evaluating parameter's values optimized. The functionality of job schedule selected may be opted manually by experts or automatically depending on characteristics of CC environment.

- *Dataset preparation*: to evaluate performance of the techniques proposed and comparative analysis, the two existing datasets synthetic and HCSP [16] are utilized. This benchmark dataset contains various types of VM instances and jobs, and uniformly distributed technique is followed to produce numerous dataset instances where the attributes of instances are described in the form of heterogeneity of resources and jobs, and consistency. The attributes are represented in the form of C-JR [16], which are uniformly distributed, where 'C' is type of consistency having values as Consistent Fully (CF), Consistent Partly (CP) and Inconsistent (IC). The fields 'J' and 'R' describes job and resource heterogeneities respectively, where higher and lower values are taken by both job and resource heterogeneities. In our result illustrations, suppose J=1 means that higher value is taken for the job and J=0 means lower value for job. Similarly, suppose R=1 means that higher value is taken for the resource and R=0 means lower value for resource. The Figure 3 gives detailed summary of attribute values of instances of datasets. The proposed research work is concentrated on instances which are produced uniformly. The HCSP (Heterogeneous Computing Scheduling Problem) dataset built based on running time for computational matrix *k* VMs and *n* jobs, which take in to consideration entire heterogeneities and consistencies of jobs as well the VMs. The HCPS data instances utilized widely for evaluation of issues related to scheduling of jobs [11, 25].

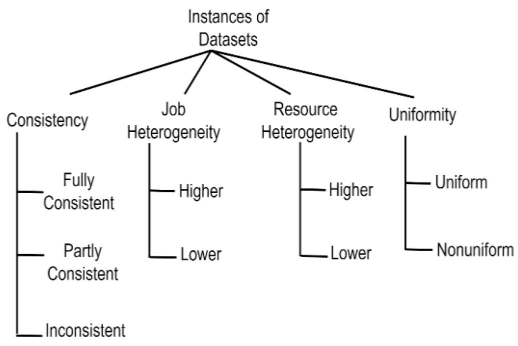

Figure 3. Data instance attributes.

For the evaluation of our proposed mechanism, we utilized synthetic and HCSP data sets, which are produced by making use of pseudo random numbers, and generated 128, 256, 512, 1024, 2048 and 4096 jobs that are heterogenous and random. As per the benchmark, the structures such as (128x4096), (256x8192), (512x16), (1024 x 32) and (2048x64) are followed, in which 128, 256, 512, 1024 and 2048 are count of jobs to be applied to 4096, 8192, 16, 32, and 64 VMs respectively.

The produced jobs are described with the notation C-JR as discussed above in this section.

For the evaluating purposes we concentrated on two levels of heterogeneities higher and lower for jobs, and cloud consistencies such as CF, CP and IC.

In the methodology adapted, firstly we opt for the instances in the benchmark data sets, then perform execution of jobs by applying proposed scheduling strategy, and lastly, calculate the parameters of evaluation such as utilization of energy and cost.

4. Results and Discussions

The setting up of experimental environment and the discussions on results obtained are described in this section. The cloud simulations software CloudSim [6] is utilized along with synthetic benchmark datasets for evaluating proposed technique. CloudSim is majorly utilized simulating tool for evaluation of optimization mechanisms, which simulates jobs, VMs and data centers, supporting strategies for scheduling the jobs and various sorts of models for energy utilization with dynamic workloads. The suggestive data center for simulating IaaS as our cloud model, provides sixteen PMs in two categories, every PM containing four VMs having three varieties of structures.

The specifications considered for cloud simulated data centers includes 2GB Primary memory, 1TB HDD, Time, 10gbps bandwidth and Xen VMM. Three types of VMs considered includes small (500 MIPS, 6 PE, 2500 capacity), medium (1000 MIPS, 9 PE, 8000 capacity) and large (1500 MIPS, 25 PE, 35000 capacity). Two types of PMs utilized are Host-I (Intel Core 2 Extreme X6800 Processor, 4 PE, 28500 MIPS) and Host-II (Intel Core i7 Extreme 3960X Processor, 8 PE, 181400 MIPS).

To assess the effectiveness of the dual-objective scheduling model, the researchers conducted a comparative analysis with several existing approaches. The evaluated techniques included Whale, ABC, PSO, and MDVMA (multi-depot vehicle routing problem with multiple ants). These techniques were chosen as benchmarks due to their established reputation and prior application in similar problem domains.

Energy utilization refers to the amount of energy consumed by the cloud infrastructure to perform various tasks and run workloads. We have taken the workloads as number of jobs staring from 5 and increasing in steps of 5 till 70 and recorded the utilization of energy to compare with the various existing approaches. Whereas, the cost of job scheduling encompasses the financial expenses associated with managing and executing tasks within a cloud environment.

We have taken the workloads for computing cost as number of jobs staring from 50 and increasing in steps of 50 till 700, recorded the cost and compared with various existing approaches.

Table 1. Synthetic datasets-based comparisons of utilization of Energy with existing techniques.

Number of jobs	**Utilization of energy by various approaches (kwh)**				
	Whale [28]	**ABC [17]**	**PSO [31]**	**MDVMA [2]**	**Proposed technique**
5	3.5	4.01	3.01	2.51	1.98
10	5.25	6.91	4.61	3.72	3.28
15	7.01	7.81	6.02	4.83	4.49
20	7.72	9.01	7.01	6.02	5.71
25	9.1	10.01	8.41	7.14	6.63
30	10.5	12.00	9.01	7.71	7.24
40	13.11	14.10	11.41	9.72	8.68
50	15.55	17.09	13.63	11.71	10.22
60	18.33	19.59	16.07	14.51	11.85
70	20.94	21.60	18.51	16.54	13.60
Overall utilization of energy	111.01	122.13	97.69	84.41	73.68
Average reduction in energy utilization using proposed technique	**33.63%**	**39.67%**	**24.58%**	**12.71%**	

The performance evaluation of the proposed technique is done by comparing utilization of energy and cost with existing four techniques such as Whale [28], ABC [17], PSO [31] and MDVMA [2] approaches. Initially the experimentation is done by making use of synthetic datasets and the results of utilization of energy for the specific number of jobs with numerous amounts of VMs are recorded in Table 1.

The simulated results ensure that energy utilization using proposed approach with respect to considered VMs are much lower compared to techniques existing. The Table 1 also shows the overall utilization of energy for set of workloads given in cloud data center and proves that the utilization of energy by proposed technique is minimized by 33.63%, 39.67%, 24.58%, and 12.71% over Whale, ABC, PSO and MDVMA respectively. The graphical comparative analysis of performance in terms of utilization of energy of proposed techniques is depicted in Figure 4. The produced results demonstrate that utilization of energy of proposed technique is at every step of amount of VMs is minimal compared to available methodologies for scheduling the jobs at data center of cloud.

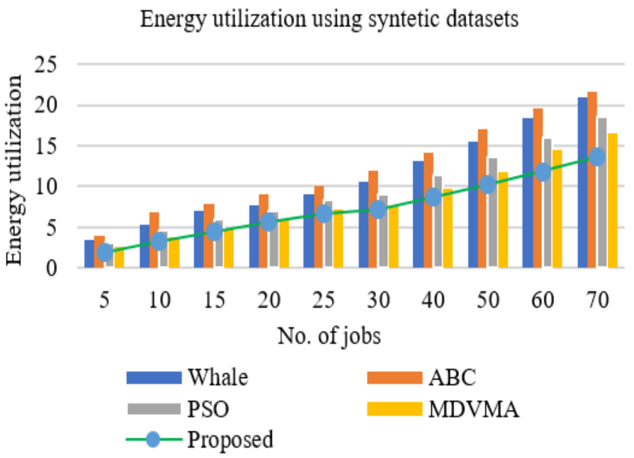

Figure 4. Comparison of energy utilization of proposed approach with existing approaches using synthetic datasets.

This research work also done the evaluation of proposed technique in terms of cost and analysed the performance by making use of synthetic datasets and the results of cost as a parameter for the specific number of jobs with numerous amounts of VMs are recorded in Table 2.

Table 2. Synthetic datasets-based comparisons of cost with existing techniques.

Number of jobs	Cost by various approaches				
	Whale [28]	ABC [17]	PSO [31]	MDVMA [2]	Proposed technique
50	98	163	79	63	57
100	175	225	150	112	104
150	263	313	238	200	188
200	351	401	326	288	234
250	439	489	414	376	326
300	527	577	502	464	412
400	703	753	678	640	593
500	879	929	854	816	763
600	1117	1167	1092	1054	924
700	1363	1413	1290	1255	1120
Overall cost	5915	6460	5623	5268	4721
Average reduction in cost using proposed technique	**20.19%**	**26.92%**	**16.05%**	**10.39%**	

The simulated results ensure that cost using proposed approach with respect to considered VMs are much lower compared to techniques existing. The Table 2 also shows the overall cost for set of workloads given in cloud data center and proves that the cost by proposed technique is minimized by 20.19%, 26.92%, 16.05%, and 10.39% over Whale, ABC, PSO, and MDVMA respectively. The graphical comparative analysis of performance in terms of cost of proposed techniques is depicted in Figure 5. The produced results demonstrate that cost of proposed technique is at every step of amount of VMs is minimal compared to available methodologies for scheduling the jobs at data center of cloud.

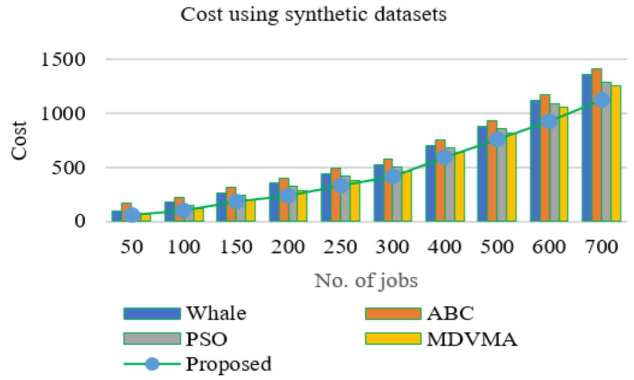

Figure 5. Comparison of Cost of proposed approach with Existing approaches using Synthetic Datasets.

The second benchmark dataset considered in this experimentation is HCSP dataset.

We have done the evaluation of proposed technique for utilization of energy and cost using HCSP datasets, analysed and compared the results with identified heuristic techniques such as Whale [28], ABC [17], PSO [31] and MDVMA [2] methods.

Various instances of dataset HCSP is utilized in the experimentation. The Table 3 records the results of using HCSP datasets for comparing Utilization of Energy with existing techniques.

Table 3. Using HCSP datasets for comparing Utilization of Energy with existing techniques.

Instances of datasets	Utilization of energy by various approaches (kwh)				
	Whale [28]	ABC [17]	PSO [31]	MDVMA [2]	Proposed technique
CF-11	55.00	60.00	58.00	43.00	36.50
CF-10	56.00	65.00	55.00	45.00	42.50
CF-01	68.00	62.00	53.00	36.00	32.00
CF-00	63.00	61.00	51.00	43.00	40.00
CP-11	68.00	75.00	66.00	52.00	48.00
CP-10	73.00	77.00	68.00	58.00	50.00
CP-01	68.00	78.00	65.00	55.00	49.50
CP-00	69.00	74.00	63.00	60.00	52.00
IC-11	78.00	85.00	75.00	65.00	55.00
IC-10	80.00	82.00	78.00	60.00	52.50
IC-01	75.00	86.00	76.00	72.00	62.00
IC-00	82.00	87.00	71.00	62.00	52.00
Overall utilization of energy	835.00	892.00	779.00	651.00	572.00
Average reduction in utilization of energy using proposed technique	**31.50%**	**35.87%**	**36.57%**	**12.14%**	

All the instances are uniformly distributed and it can be CF, CP, and IC, with values for job and resource heterogeneities. As seen in Table 3, values 1 and 0 are used for mentioning higher and lower value for job and resource heterogeneities. The entry CF-11 in Table 3 illustrates that the data instance is CF with high job heterogeneity and high resource heterogeneity.

The entry CP-01 illustrates that the data instance is CP with low job heterogeneity and high resource heterogeneity. Likewise, entry IC-00 illustrates that data instance is IC with low job heterogeneity and low resource heterogeneity.

The Table 3 also shows the overall utilization of energy for set of workloads given in cloud data center and proves that the utilization of energy by proposed technique is minimized by 31.50%, 35.87%, 36.57%, and 12.14% over Whale, ABC, PSO, and MDVMA respectively. The graphical comparative analysis of performance in terms of utilization of energy of proposed techniques is depicted in Figure 6.

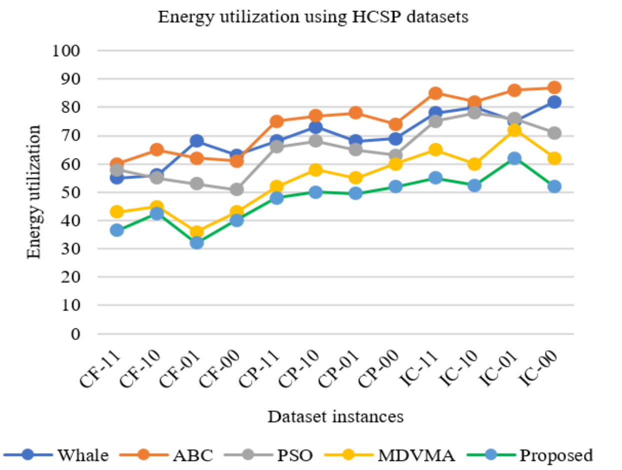

Figure 6. Comparison of energy utilization of proposed approach with existing approaches using HCSP datasets.

The produced results demonstrate that utilization of energy of proposed technique is at every step of amount of VMs is minimal compared to available methodologies for scheduling the jobs at data center of cloud.

The evaluation of proposed technique is also done in terms of cost and analysed the performance by making use of HCSP datasets and the results of cost as a parameter for the specific number of jobs with numerous amounts of VMs are recorded in Table 4. The simulated results ensure that cost using proposed approach with respect to considered VMs are much lower compared to techniques existing.

Table 4. Using HCSP datasets for comparing Cost with existing techniques.

Instances of datasets	Cost by various approaches (kwh)				
	Whale [28]	ABC [17]	PSO [31]	MDVMA [2]	Proposed technique
CF-11	65	54	43	60	40
CF-10	63	55	46	65	42.5
CF-01	62	58	35	62	30
CF-00	61	55	44	61	42
CP-11	67	68	51	75	48.5
CP-10	68	72	58	78	52
CP-01	65.5	69	55	78.4	50
CP-00	63	69.5	59	74	56.5
IC-11	74	78	65	85	62.5
IC-10	76	80	63	84	60
IC-01	75	77	71	85.5	65
IC-00	72	81	65	85.5	62.5
Overall cost	811.5	816.5	655	893.4	611.5
Average reduction in cost using proposed technique	**24.65%**	**25.11%**	**6.64%**	**31.55%**	

All the instances are uniformly distributed and it can be CF, CP and IC, with values for job and resource heterogeneities.

As seen in Table 4, values 1 and 0 are used for mentioning higher and lower value for job and resource heterogeneities. The entry CF-11 in Table 3 illustrates that the data instance is CF with high job heterogeneity and high resource heterogeneity. The Table 4 also shows the overall cost for set of workloads given in cloud data center and proves that the cost by proposed technique is minimized by 24.65%, 25.11%, 6.64%, and 31.55% over Whale, ABC, PSO, and MDVMA respectively. The graphical comparative analysis of performance in terms of cost of proposed techniques is depicted in Figure 7. The produced results demonstrate that cost of proposed technique is at every step of amount of VMs is minimal compared to available methodologies for scheduling the jobs at data center of cloud.

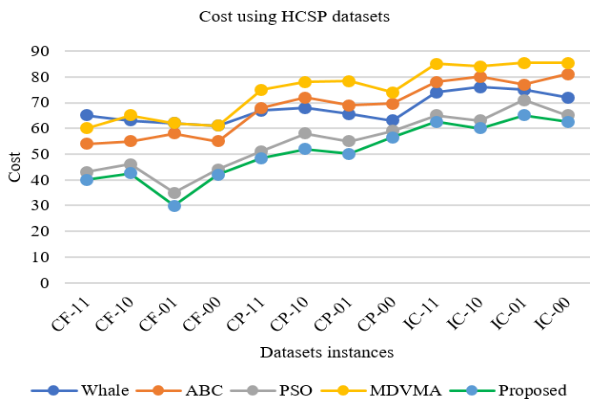

Figure 7. Comparison of cost of proposed approach with existing approaches using HCSP datasets.

Hence, it could be concluded that minimization of cost and energy utilization will impact on improving income of providers of cloud services by making use of proposed technique for scheduling the jobs.

4.1. Limitation of the Work

While the proposed work offers significant contributions to addressing the challenges in CC environments, there are some limitations that should be acknowledged. These limitations include:

- *Simplified* assumptions: the proposed framework may rely on certain assumptions to simplify the problem and make it computationally tractable. These assumptions may not capture the full complexity and variability of real-world CC environments, potentially limiting the generalizability of the framework.
- *Limited resource diversity*: the proposed framework may assume homogeneity in resource characteristics, such as processing power, memory capacity, and network bandwidth. However, in real cloud environments, the resources can be heterogeneous, varying in performance and capabilities. Failure to consider such diversity may limit the accuracy and effectiveness of the allocation and scheduling decisions.
- *Dependency on benchmark datasets*: the

performance evaluation of the proposed framework relies on benchmark datasets, which may not fully capture the diverse real-world scenarios encountered in CC environments. The limitations of the datasets may impact the generalizability and validity of the results obtained.

4.2. Future Research

The proposed work opens up several avenues for future research and potential improvements. Some areas worth exploring include:

- *Dynamic and real-time adaptation*: enhance the framework to dynamically adapt to real-time changes in job arrivals, resource availability, and workload fluctuations. Incorporate mechanisms to handle dynamic allocation and scheduling decisions in response to varying conditions, ensuring responsiveness and adaptability in dynamic cloud environments.
- *Multi-objective optimization*: extend the framework to support a more comprehensive multi-objective optimization approach. Consider additional objectives such as energy efficiency, reliability, and quality of service, allowing users to customize the trade-offs between multiple objectives based on their specific requirements and preferences.
- *Reinforcement learning and AI techniques*: explore the application of reinforcement learning and other artificial intelligence techniques to optimize the allocation and scheduling decisions. Design intelligent agents that learn from past experiences and adapt their decision-making processes to achieve improved job scheduling efficiency and resource utilization.
- *Multi-tenancy support*: extend the framework to support multi-tenancy, enabling efficient allocation and scheduling for multiple users or tenants sharing the same cloud infrastructure. Consider resource isolation, fairness, and security aspects to ensure efficient and secure resource allocation among different tenants.

By focusing on these areas for future research and improvements, the proposed work can advance the field of VM allocation and job scheduling in CC, leading to more efficient resource utilization, enhanced performance, and improved user satisfaction in cloud-based applications.

4.3. Scalability of the Proposed Technique

The scalability of the proposed technique is an important aspect to consider. Scalability refers to the ability of the technique to handle increasing workloads and larger CC environments without significant degradation in performance. Here, we discuss the scalability of the proposed technique in the following aspects:

- *Number of VMs*: the scalability of the technique is evaluated based on its ability to efficiently allocate jobs and schedule tasks across a growing number of VMs. As the number of VMs increases, the technique should demonstrate the capability to handle the increased computational load and make optimal allocation decisions in a timely manner. The efficiency of the algorithm and the optimization approach employed play a crucial role in maintaining scalability.
- *Number of jobs and workload size*: scalability is also assessed by evaluating the performance of the proposed technique as the number of jobs and workload size increases. The technique should be able to handle larger workloads without experiencing a significant increase in execution time or resource consumption. The allocation and scheduling decisions should be made efficiently, ensuring that the system can accommodate the growing number of jobs and effectively utilize available resources.
- *Resource utilization and load balancing*: the scalability of the proposed technique is linked to its ability to effectively utilize resources in larger cloud environments. The technique should demonstrate efficient load balancing capabilities, ensuring that resources are optimally allocated across VMs to avoid resource bottlenecks and maximize resource utilization. As the number of VMs and workload size grows, the technique should adapt and distribute the workload evenly to maintain scalability.

Addressing the scalability challenges is crucial for the proposed technique to be applicable in practical CC environments. By ensuring that the technique can handle increasing workloads, larger numbers of VMs, and diverse resource demands efficiently, the scalability of the proposed dual-objective framework can be established, providing robustness and effectiveness in real-world scenarios.

4.4. The Computational Complexity Involved in Implementing the Proposed Technique

This can vary depending on several factors, including the algorithmic approach chosen, the size of the problem, and the efficiency of the implementation. Here are some considerations regarding computational overhead and complexity:

- *Mathematical model and optimization algorithm*: the complexity of the mathematical model used in the framework and the optimization algorithm employed significantly influence the computational overhead.
- *Problem size and dimensionality*: the computational overhead typically increases with the size and dimensionality of the problem. As the number of VMs, jobs, and associated parameters grows, the complexity of the optimization problem increases. Large-scale problems with a high number of

variables and constraints may require more computational resources and time to obtain optimal or near-optimal solutions.

5. Conclusions

Work done in this paper presents a dual-objective approach for allocation of VM with improved scheduling of job in the environment of CC. The dual-objective model optimizes two parameters utilization of energy and cost by providing solutions to problem of scheduling a job. The simulation results demonstrated that dual-objective approach placed better compared to existing methodologies, including Whale, ABC, PSO and MDVMA techniques, by making use of synthetic and HCSP datasets, in terms of optimized cost and utilization of energy. The dual-objective approach performed better than existing techniques by minimizing utilization of energy by 33.63%, 39.67%, 24.58% and 12.71%; and cost by 20.19%, 26.92%, 16.05%, and 10.39% over Whale, ABC, PSO and MDVMA respectively, utilizing synthetic datasets. A remarkable minimization of cost and utilization of energy also demonstrated utilizing HCSP datasets, considering various constraints such as consistencies, uniformities, and heterogeneities of job and resources.

References

[1] AlJahdali H., Albatli A., Garraghan P., Townend P., Lau L., and Xu J., "Multi-Tenancy in Cloud Computing," *in Proceedings of the IEEE 8th International Symposium on Service Oriented System Engineering*, Oxford, pp. 344-351, 2014. DOI:10.1109/SOSE.2014.50

[2] Alsadie D., "A Metaheuristic Framework for Dynamic Virtual Machine Allocation with Optimized Task Scheduling in Cloud Data Centers," *IEEE Access*, vol. 4, pp. 74218-74233, 2021. DOI:10.1109/ACCESS.2021.3077901

[3] Ayyapazham R. and Velautham K., "Proficient Decision Making on Virtual Machine Creation in IaaS Cloud Environment," *The International Arab Journal of Information Technology*, vol. 14, no. 3, pp. 314-323, 2017. https://www.iajit.org/PDF/Vol%2014%2C%20No.%203/7679.pdf

[4] Basthikodi M. and Ahmad W., "Parallel Algorithm Performance Analysis Using OpenMP for Multicore Machines," *International Journal of Advanced Computer Technology*, vol. 4, no. 5, pp. 28-32, 2015. https://www.ijact.org/ijactold/volume4issue5/IJ0450005.pdf

[5] Basthikodi M., Prabhu A., and Bekal A., "Performance Analysis of Network Attack Detection Framework using Machine Learning," *Sparkinglight Transactions on Artificial Intelligence and Quantum Computing*, vol. 1, no. 1, pp. 11-22, 2021. file:///C:/Users/user/Downloads/staiqc-paper2.pdf

[6] Calheiros R., Ranjan R., Beloglazov A., De Rose C., and Buyya R., "Cloudsim: A ToolKit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms," *Software: Practice and Experience*, vol. 41, no. 1, pp. 23-50, 2011. https://doi.org/10.1002/spe.995

[7] Deb K., Pratap A., Agarwal S., and Meyarivan T., "A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182-197, 2002. DOI:10.1109/4235.996017

[8] Ebadifard F. and Babamir S., "Optimizing Multi Objective Based Workflow Scheduling in Cloud Computing Using Black Hole Algorithm," *in Proceedings of the 3rd International Conference on Web Research*, Tehran, pp. 102-108, 2017. DOI:10.1109/ICWR.2017.7959313

[9] Elaziz M., Xiong S., Jayasena K., and Li L., "Task Scheduling in Cloud Computing Based on Hybrid Moth Search Algorithm and Differential Evolution," *Knowledge-Based Systems*, vol. 169, pp. 39-52, 2019. https://doi.org/10.1016/j.knosys.2019.01.023

[10] Fazio M., Ranjan R., Girolami M., Taheri J., Dustdar S., and Villari M., "A Note on the Convergence of IoT, Edge, and Cloud Computing in Smart Cities," *IEEE Cloud Computing*, vol. 5, no. 5, pp. 22-24, 2018. DOI:10.1109/MCC.2018.053711663

[11] Gajera V., Shubham., Gupta R., and Jana P., "An Effective Multi-Objective Task Scheduling Algorithm Using Min-Max Normalization in Cloud Computing," *in Proceedings of the 2nd International Conference on Applied and Theoretical Computing and Communication Technology*, Bangalore, pp. 812-816, 2016. DOI:10.1109/ICATCCT.2016.7912111

[12] Hameed A., Khoshkbarforoushha A., Ranjan R., and Jayaraman P., "A Survey and Taxonomy on Energy Efficient Resource Allocation Techniques for Cloud Computing Systems," *Computing*, vol. 98, no. 7, pp. 751-774, 2016. DOI:10.1007/s00607-014-0407-8

[13] Hazra D., Roy A., Midya S., and Majumder K., *Smart Innovation, Systems and Technologies*, Springer, 2018. https://doi.org/10.1007/978-981-10-5544-7_19

[14] Holagundi N., Ashwathsetty G., and Basthikodi M., "Algorithm Fuzzy Scheduling (AFS) for Realtime Jobs on Microprocessor Systems," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 25, no. 3, pp. 1308-1319, 2022.

https://ijeecs.iaescore.com/index.php/IJEECS/issue/view/588

[15] Hosseinzadeh M., Ghafour M., Hama H., Vo B., and Khoshnevis A., "Multi-Objective Task and Workflow Scheduling Approaches in Cloud Computing: A Comprehensive Review," *Journal of Grid Computing*, vol. 18, pp. 327–356, 2020. https://doi.org/10.1007/s10723-020-09533

[16] Hussain A. and Aleem M., "GoCJ: Google Cloud Jobs Dataset for Distributed and Cloud Computing Infrastructures," *Data*, vol. 3, no. 4, pp. 1-12, 2018. https://doi.org/10.3390/data3040038

[17] Jena R., "Task Scheduling in Cloud Environment: A Multi-Objective ABC Framework," *Journal of Information and Optimization Sciences*, vol. 38, no. 1, pp. 1-19, 2017. https://doi.org/10.1080/02522667.2016.1250460

[18] Junaid M., Sohail A., Ahmed A., and Baz A., "A Hybrid Model for Load Balancing in Cloud Using File Type Formatting," *IEEE Access*, vol. 8, pp. 118135-118155, 2020. DOI:10.1109/ACCESS.2020.3003825

[19] Khattar N., Sidhu J., and Singh J., "Toward Energy-Efficient Cloud Computing: A Survey of Dynamic Power Management and Heuristics-Based Optimization Techniques," *The Journal of Supercomputing*, vol. 75, pp. 4750-4810, 2019. https://doi.org/10.1007/s11227-019-02764-2

[20] Kuang P., Guo W., Xu X., Li H., Tian W., and BuyyaR., "Analysing Energy-Efficiency of two Scheduling Policies in Compute-Intensive Applications on Cloud," *IEEE Access*, vol. 6, pp. 45515-45526, 2018. DOI:10.1109/ACCESS.2018.2861462

[21] Lagana D., Mastroianni C., Meo M., and Renga D., "Reducing the Operational Cost of Cloud Data Centers through Renewable Energy," *Algorithms*, vol. 11, no. 10, pp. 1-21, 2018. https://doi.org/10.3390/a11100145

[22] Masdari M. and Zangakani M., "Green Cloud Computing Using Proactive Virtual Machine Placement: Challenges and Issues," *Journal of Grid Computing*, vol. 18, pp. 727-759, 2020. https://doi.org/10.1007/s10723-019-09489-9

[23] Masdari M., Salehi F., Jalali M., and Bidaki M., "A Survey of PSO Based Scheduling Algorithms in Cloud Computing," *Journal of Network and Systems Management*, vol. 25, pp. 122-158, 2017. https://doi.org/10.1007/s10922-016-9385-9

[24] Omkar S., Khandelwal R., Ananth T., Naik G., and Gopalakrishnan S., "Quantum Behaved Particle Swarm Optimization (QPSO) for Multi-Objective Design Optimization of Composite Structures," *Expert Systems with Applications*, vol. 36, no. 8, pp. 11312-11322, 2009. https://doi.org/10.1016/j.eswa.2009.03.006

[25] Panda S. and Jana P., "A Multi-Objective Task Scheduling Algorithm for Heterogeneous Multi-Cloud Environment," *in Proceedings of the International Conference on Electronic Design, Computer Networks and Automated Verification*, Shillong, pp. 82-87, 2015. DOI:10.1109/EDCAV.2015.7060544

[26] Patel N. and Patel H., "Energy Efficient Strategy for Placement of Virtual Machines Selected from Underloaded Servers in Compute Cloud," *Journal of King Saud University-Computer and Information Sciences*, vol. 32, no. 6, pp. 700-708, 2020. https://doi.org/10.1016/j.jksuci.2017.11.003

[27] Pietri I., Chronis Y., and Ioannidis Y., "Multi-Objective Optimization of Scheduling Dataflows on Heterogeneous Cloud Resources," *in Proceedings of the IEEE International Conference on Big Data*, Boston, pp. 361-368, 2017. DOI:10.1109/BigData.2017.8257946

[28] Reddy G. and Kumar S., "Multi Objective Task Scheduling Algorithm for Cloud Computing Using Whale Optimization Technique," *in Proceedings of the 3rd International Conference of Smart and Innovative Trends in Next Generation Computing Technologies*, Dehradun, pp. 286-297, 2017. https://doi.org/10.1007/978-981-10-8657-1_22

[29] Sanaj M. and Prathap P., "An Efficient Approach to the Map-Reduce Framework and Genetic Algorithm-Based Whale Optimization Algorithm for Task Scheduling in Cloud Computing Environment," *Materials Today: Proceedings*, vol. 37, pp. 3199-3208, 2021. https://doi.org/10.1016/j.matpr.2020.09.064

[30] Sreenu K. and Sreelatha M., "W-Scheduler: Whale Optimization for Task Scheduling in Cloud Computing," *Cluster Computing*, vol. 22, pp. 1087-1098, 2019. https://doi.org/10.1007/s10586-017-1055-5

[31] Suresh S., Sujit P., and Rao A., "Particle Swarm Optimization Approach for Multi-Objective Composite Box-Beam Design," *Composite Structures*, vol. 81, no. 4, pp. 598-605, 2007. https://doi.org/10.1016/j.compstruct.2006.10.008

[32] Usman M., Ismail A., Abdul-Salaam G., and Chizari H., "Energy-Efficient Nature-Inspired Techniques in Cloud Computing Datacenters," *Telecommunication Systems*, vol. 71, pp. 275-302 2019. https://doi.org/10.1007/s11235-019-00549-9

[33] Varghese B. and Buyya R., "Next Generation Cloud Computing: New Trends and Research Directions," *Future Generation Computing System*, vol. 79, pp. 849-861, 2018. https://doi.org/10.1016/j.future.2017.09.020

[34] Yang L., Cao J., Liang G., and Han X., "Cost Aware Service Placement and Load Dispatching

in Mobile Cloud Systems," *IEEE Transactions on Computers*, vol. 65, no. 5, pp. 1440-1452, 2016. DOI:10.1109/TC.2015.2435781

[35] Yao Y., Wang Z., and Zhou P., "Privacy-Preserving and Energy Efficient Task Offloading for Collaborative Mobile Computing in IoT: An ADMM Approach," *Computers and Security*, vol. 96, pp. 101886, 2020. https://doi.org/10.1016/j.cose.2020.101886

Sandeep Sutar is pursuing Ph.D. from Department of Computer Science and Engineering, Don Bosco Institute of Technology, Bangalore. He has 18 years of teaching experience. Currently he is working in the Department of Computer Science and Engineering at Annasaheb Dange College of Engineering and Technology, Ashta, India. His area of interest includes Cloud Computing, IOT and Artificial Intelligence.

Manjunathswamy Byranahallieraiah obtained his Doctoral Degree in Computer Science and Engineering from UVCE, Bangalore University, Bangaluru. Currently he is working in Department of Computer Science and Engineering, Don Bosco Institute of Technology, Bangalore. His area of interest includes Image Processing, Signal Processing, and Network Security, Cloud Computing, IOT and Data Science.

Kumaraswamy Shivashankaraiah is currently working as an Assistant Professor in the Department of Computer Science and Engineering, University Visveswaraya College of Engineering, Bengaluru. He received his Ph.D in Computer Science and Engineering from UVCE, Bangalore University, Bangalore. His research interest is in the area of Data mining, Web mining, Semantic web and cloud computing.