IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 15 July 2022, accepted 31 July 2022, date of publication 3 August 2022, date of current version 9 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3196240

RESEARCH ARTICLE

SLA-Aware and Energy-Efficient Virtual Machine Placement and Consolidation in Heterogeneous DVFS Enabled Cloud Datacenter

BADIEH NIKZAD[1], BEHNAM BARZEGAR[1], AND HOMAYUN MOTAMENI[2]
[1]Department of Computer Engineering, Babol Branch, Islamic Azad University, Babol 47471-37381, Iran
[2]Department of Computer Engineering, Sari Branch, Islamic Azad University, Sari 48161-19318, Iran

Corresponding author: Behnam Barzegar (Barzegar.behnam@yahoo.com)

ABSTRACT Increasing demand for computational resource as services over the internet has led to the expansion of datacenter infrastructures. Thus, datacenter authorities are striving to adopt optimal power usage schemes to minimize costs, emissions and Service Level Agreement (SLA) violations in their task scheduling for heterogeneous computation centers. One of the most effective strategies to reduce datacenter energy consumption is to maximize the utilization of physical machines and shut down the idle ones. This can be realized through two main algorithms, namely virtual machine placement and virtual machine consolidation. The VM placement method is a dynamic process to put these virtual devices on physical machines. The consolidation technique, however, tries to improve physical machine efficiency through grouping and live migration of dispersed virtual machines on lower number of active physical machine. In this paper, a novel approach is proposed for improving the physical machine efficiency. The approach employs heuristics and meta-heuristic algorithms with eight performance criteria and is implemented on small to medium scale data centers using simulated cloud module. The results indicates that the proposed method showed up to 10.3%, 5.3%, and 12.5% the more significant efficiency rather best previous algorithms, respectively, in terms of the energy consumption, number of SLA violation and number of VMs migration.

INDEX TERMS Green data centers, service level agreement, virtual machine placement, virtual machine consolidation.

I. INTRODUCTION

During the past few years, computation technology has attempted to adapt to the ever-increasing demand of high-level computational services and equipment [1]. This evolution has brought about novel computational ideas like cloud computing. These systems have become a well-adopted paradigm for hosting a multitude of computational service providers. These services may include Platform as a Service (PaaS), Software as a Service (SaaS) and Infrastructure as a Service (IaaS). The computing systems thus access the computational resources and deliver pay-per-demand services to end users. Hence, a main feature of cloud computing, and infrastructure-based services in particular, is the incorporation of virtualization technology [2].

The associate editor coordinating the review of this manuscript and approving it for publication was Fabrizio Messina.

The issue of tremendous energy consumption by green data centers caused by the increasing computational load has become a critical research topic. As reported by Gartner in 2013, the average energy usage of a data center equals that of around 25,000 households [3]. Also, another survey revealed the federal could computing energy consumption in the US to be around 100 million MWh with an associated cost making up about 75% of the entire operating costs [4]. The estimated annual energy consumption of US data centers in 2020 was 140 million MWh with a cost of $13 billion billed for cloud consumers [5].

Thus, a significant amount of electricity cost is incurred for these high-tech infrastructures while CO_2 emissions also escalate. As collateral, the high energy usage level will imply higher cooling demand and costs. The other disvantage is the wear and tear of computational devices through high temperatures influencing their availability and reliability while causing high SLA (service level agreement) violations. SLA is defined as the service level expected from the facility vendor [6].

The focus of green cloud computing is on designing data centers that reduce energy consumption by increasing the efficiency of resources and physical machines. One of the main approaches used for optimizing physical machine utilization and reduction of energy consumption in cloud data centers is to optimize the utilization of physical machines (PMs) and switch the idle PMs to sleep mode or scale down PM's operating frequency to the lowest for DVFS-enabled Cloud datacenters. Two main classes of algorithms have been proposed to realize this approach: VM placement algorithms and VM consolidation algorithms [7].

VM placement consists of the dynamic mapping of VMs onto PMs in cloud datacenters, in a way to optimize resource utilization [8]. Another potential way to enhance energy efficiency is the VM workload consolidation strategy in which energy consumption is minimized by consolidating higher workload on lower number of PMs. VMs hosted on lightly loaded PMs are live migrated and group-dispersed on a minimal number of active PMs While the idle PMs are turned off. Optimization works on VM placement and consolidation must deal with a diverse set of challenges. First and for most, these problems are of NP-hard type requiring tremendous amount of time and resources [5]. The second challenge involves taking care of the system performance while attempting to reduce energy consumption. By shutting down some of the physical machines, some VMs may face resource shortage especially during peak periods. As a consequence, reliability and/or availability of the system is overshadowed leading to lower-than-expected QOS levels. Effective VM consolidation stipulates performance maintenance and SLA violation restriction when dealing with unpredictable computational demand [9]. Thus, the significance of establishing a compromise between energy usage and performance metrics in optimal resource management schemes is readily inferred. This task is indeed among the main challenges to be handled by providers of cloud services. Given the NP nature of the problem and multiple contradictory objectives, the proposed approach employs the multi-objective ant colony optimization which yields solutions within Pareto front [5]. This paper proposes a two-phase energy- and SLA-aware multi-objective VM placement and consolidation approach which employs the DVFS technique aimed to achieve a trade-off between energy consumption and SLA. In the first phase, the MRAT-MACO VM placement algorithm is developed which seeks to find optimal VM placement solutions in order to minimize the total energy consumption, CPU wastage and communication energy cost in a DVFS-enabled cloud datacenter. The live VM migration which transfers a running VM from a PM to another with no interruption in services can be performed in the second phase i.e. MRAT-MACO VM consolidation algorithm.

For the purpose of validating our proposed approach, test beds generated by Cloudsim tool are utilized with diverse series of configurations. In addition, the number of loads and resources are varied to get a comprehensive analysis of data center performance. The generated test beds are also used to simulate a series of other single-objective methods including DVFS, LR, FFD and ST as well as multi-objective methods such as MGA and MACO-Feller to draw a performance comparison against the proposed approach. The performance assessment takes account of different metrics including energy usage, percentage of resource wasting, saved energy, cost of communication energy, level of VM migration, SLA violation instances, execution period and the number of active PMs. The results indicates that the proposed method showed up to 10.3%, 5.3%, and 12.5% the more significant efficiency rather best previous algorithms, respectively, in terms of the energy consumption, number of SLA violation and number of VMs migration.

Thus, the proposed approach in this paper offers the following contributions:

- A multi-objective QOS-aware and energy-aware approach for cloud resource management is proposed aimed at efficient energy consumption while achieving a high QOS and SLA fulfillment.
- A thorough performance comparison is made between the proposed approaches against six other single- and multi-objective methods considering eight distinctive metrics. This will help gain insight into the impacts of considering each objective on the resulting energy efficiency and SLA compliance. Further, a comparative assessment is carried out between heuristic and meta-heuristic methods and their performance in terms of VM placement and consolidation efficiency.

The remaining parts of this paper are organized in five sections. Section 2 gives a review of the literature and analyzes the encountered limitations in the existing methods. Section 3 elaborates on the proposed optimization scheme. Simulation results of the proposed method and the related performance analyses are presented in section 4. Finally, section 5 draws the main conclusions.

II. RELATED WORK

This section discusses the main existing algorithms currently employed for energy-efficient resource allocation based on VM placement and consolidation within cloud data centers.

A. VM PLACEMENT STRATEGIES

VM placement is a dynamic method to map virtual machines to physical machines enabling them to share resources in an efficient way. Feller *et al.* [10] proposed an ant colony optimization strategy in which VM placement is considered as a multi-dimensional grouping problem. The primary objective is to put all items within the minimum space. The approach is weighed against the greedy approach of First Fit Decreasing (FFD) algorithm. The results indicate superior performance of ant colony approach compared to FFD in terms of lowering the consumed energy.

Gao *et al.* [11] employed multi-objective ant colony algorithm for VM placement. The approach aims to yield a set of solutions which concurrently minimize resource wastage and power consumption.

The proposed algorithm is weighed against multi-objective genetic algorithm, bin-packing approach and max–min ant system (MMAS) method. The results indicate that their proposed algorithm outperforms the mentioned methods.

Sarma *et al.* [12] used a combination of ant colony and multi-objective genetic algorithm to obtain the optimal VM placement solution. The simulation results verify that the consolidation approach yields better outcomes and further reduces costs and energy usage of the servers.

Shabeera *et al.* [13] utilized the meta heuristic algorithm of ant colony optimization for optimal allocation of virtual machines and data placement for data-intensive application in the cloud. The main goal is to relieve network traffic and bandwidth through optimal VM and data placement in physical machines in proximity. In fact, the algorithm attempts to find proximate groups of physical machines for data placement. The data are distributed among the storage elements of the selected physical machines. Based on the PM processing capacity, the required quantity of VMs is assigned to process the stored data.

Sun *et al.* [14] proposed multi-population ant colony algorithm for optimal VM placement. Their approach uses several ant colonies with strategies for exchange of population entropy data between the colonies as required for the diversity of explored solutions and their convergence. The approach is demonstrated to provide better solutions in comparison to single-population AC method and efficiently lowers resource wastage and energy consumption for high-demand virtual machine deployment.

Liu *et al.* [15] proposed multi-objective Ant Colony System (ACS) algorithm for VM placement in data centers with intense bandwidths. They attempt to obtain a set of optimal Pareto solutions to maximize communication revenues and, in the meantime, minimize the energy consumption of physical machines.

Further, a VM placement based on energy-efficient ant colony algorithm is introduced by Qin *et al.* [16] in which the number of active physical servers are minimized using energy-efficient evolutionary computing techniques. Their approach utilizes Ant Colony System (ACS) to achieve the optimal results for VM placement.

Alharbi *et al.* [17] introduced an efficient optimization method based on ant colony approach with novel exploration algorithms. The efficiency of the proposed AC method is demonstrated by application to various scales of data centers. The results are weighed against those of two other methods indicating that the proposed approach has better energy efficiency for data centers on all scales. Also, the AC method exhibits decent scalability commensurate with the problem size.

Wang *et al.* [18] use energy-aware particle swarm optimization (PSO) for VM placement in heterogeneous data centers. In their proposed approach, a virtual replacement system with lower energy consumption is introduced. The results imply an energy saving of 13 to 23 percent.

Suseela *et al.* [19] proposed a multi-objective optimization algorithm of ACO-PSO for VM placement in cloud computing. The algorithm attempts to minimize resource wastage and energy consumption and balance load on the physical servers.

Dong *et al.* [20] applied VM placement for optimal network performance in cloud data centers. The authors propose a combination of Ant colony optimization and 2-opt local search to accomplish optimization goal and attempt to reduce the total communication traffic in the cloud data center network. This is developed as a quadratic assignment problem aimed at optimizing the network link utilization.

Liu *et al.* [21] proposed a unified procedure based on ant colony system for dynamic VM consolidation and live migration in cloud computing. In this method, adequate servers are initially allocated to host VMs and are then gradually reduced. The proposed method attempts to capture feasible solutions with minimum VM migrations for any given number of servers.

Dong *et al.* [22] employed decentralized parallel genetic algorithm (DPGA) for placement of VMs and their resetting in cloud platform. In the first stage, the genetic algorithm is executed in distributed manner on several selected physical hosts in parallel. Then the algorithm continues to execute the genetic algorithm in the second stage using first-stage solutions as the initial population. The results show that DPGA can guarantee acceptable QOS/SLA for users while being more energy-efficient than other placement strategies for cloud data centers.

Wu *et al.* [23] introduced a modified genetic algorithm for VM placement in data centers. They use the server consolidation techniques based on virtualization for improved energy efficiency of both physical machines and communication networks of a data center leading to enhanced overall system performance.

Joseph *et al.* [24] also propose an approach for allocating virtual machines using genetic algorithm. The results indicate that the proposed approach is capable of lowering energy consumption and migrations.

The authors propose a multi-objective function for dynamic VM placement with live migration to minimize the resource wastage, over commitment ratio and migration energy all at the same time. Island NSGA-II optimization algorithm adopts a novel evolutionary meta-heuristic method based on an island population model to estimate the Pareto optimal set of VM placements with acceptable accuracy and diversity.

As demonstrated by the simulation results, this method outperforms related methods by reducing the migration energy [25].

In [26], authors propose a RAA-PI-NSGAII method for resource allocation algorithm using the minimum number of physical machines used and the minimum distances of resource performance and resource proportion.

In [27], the authors propose a NSGA-III method to achieve optimal multi-objective virtual machine placement (MO-VMP) based on by the minimizing the number of used physical machines and the minimum distances of resource performance and resource proportion.

In [52], authors propose two low-overhead heuristic algorithms called Global Slack Aware Quality-level Allocator (G-SLAQA) and Total Slack Aware Quality-level Allocator (T-SLAQA) for problem of scheduling a real-time application as a single DTG, where tasks may have multiple implementations designated as quality-levels, with higher quality-levels producing more accurate results and contributing to higher Quality-of-Service for the system. First, authors introduce an optimal solution using Integer Linear Programming (ILP) for a DTG with multiple quality-levels, to be executed on a heterogeneous distributed platform.

As listed in Table 1, significant research has been conducted to address various metrics and offer effective solutions.

Several issues related to resource allocation in cloud computing data centers have been discussed. Existing dynamic VMs' consolidation approaches allow the minimization of resource wastage and the reduction of energy and power consumption by switching unused PMs to idle mode. However, reducing energy consumption by means of resource consolidation may degrade the system performance and lead to SLAs' violation. Therefore, the optimal resource allocation technique should achieve a tradeoff between energy and cloud data center performance. Many resource allocation approaches focused on maximizing performance without considering energy consumption. However, even the energy-aware approaches proposed have some limitations. Indeed, if we turn off some physical servers for the purpose of saving energy, some VMs cannot receive the required resources in peak time. As a result, the system reliability and availability will be reduced and the SLA cannot be achieved. The service providers should avoid violations and keep a check while providing services to the customers. To address this issue, various researchers provide different solutions. Therefore, multi-objective optimization approaches that evaluate various parameters should be considered. Currently, the existing approaches focus on achieving high speed or high scalability, but did not address other important objectives, such as resource utilization, consolidation cost, reliability, and availability. Moreover, in order to build a practical approach that applicable in production environments, different parameters should be considered by the resource management strategy such as CPU, Memory, storage, and Network bandwidth.

B. VM CONSOLIDATION STRATEGIES

A massive amount of energy is consumed by cloud data centers. Thus, these facilities also play a big role in high CO_2 emission [28]. One potential strategy to solve this issue and optimize resource utilization in cloud data centers is to consolidate more workload on lower number of PMs or consolidate several VMs onto a PM and switch the idle PMs to sleep mode with lowest level of frequency [29]–[31]. The main feature that makes the VMC techniques interesting is live VM migration from lightly loaded PMs to comparatively higher PMs with no interruption in services [32], [33]. VM consolidation can be accomplished in different ways according to criteria, resources, objectives, and algorithmic methods [34], [35], [38], [51]. Due to the importance of VM consolidation, some research has been conducted to examine this approach to lower energy consumption in cloud data centers.

Jiang *et al.* [36] present the fast artificial bee colony based on live VM consolidation policy along with a data-intensive energy model or so-called DataABC. In this approach, a new energy evaluation model with CPU and GPU utilization rates is introduced. Also, two live VM consolidation techniques, one for VM selection and the other for VM allocation, are employed.

Mazumdar and Pranzo [39] propose a MILP mathematical formulation based on snapshot solution for server consolidation problem via live VM migration from Cloud infrastructure provider. This method aims at reducing power expenses by efficient consolidation of running server workloads and also minimizing overhead by reducing the total number of VM migrations.

Zheng *et al.* [40] propose a new solution to the virtual machine consolidated placement problem called VMPMBBO. The proposed VMPMBBO deals with virtual machine consolidated placement problem in cloud data centers and utilizes an optimization algorithm based on biogeography to optimize the virtual machine placement that minimizes both the resource wastage and the energy consumption at the same time. Extensive experiments have been conducted using synthetic data obtained from literature as well as two real datasets. The proposed method is compared with two existing multi-objective VMcP optimization techniques and is shown to have superior convergence characteristics and more computationally efficient and robust.

In addition, a normalization-based VM consolidation technique (NVMC) is proposed in [41] with online placement of VMs with the objective of minimizing energy consumption, SLA violations and VM migrations. In this approach, the overloaded hosts on a platform of virtualized cloud are identified through resource parameters. The capacity of Virtual machines and hosts are monitored to detect overloaded hosts and the cumulative available-to-total ratio (CATR) index helps identify lightly loaded machines.

Aryania *et al.* [42] propose a distributed Ant Colony Optimization System (ACOS) to save the energy consumption of cloud data centers. In their study, a new algorithm to solve the VMC problem aims to reduce the number of VM migrations, number of sleeping PMs, number of SLA violations, and reduce CCS energy consumptions.

In [43], a centralized approach is proposed based on greedy methods to solve VMC problem. The authors use the MBFD

TABLE 1. Summary of VM placement techniques.

Ref	Method	Energy & power	Server utilization	Resource wastage	Network traffic	Bandwidth usage	Revenue	load balancing	QOS	Cost	Energy migration
Feller et al [10]	ACO	✓	✓								
Gao et al [11]	MACO	✓		✓							
Kumar Sarma et al [12]	ACO & GA	✓	✓								
Shabeera et al [13]	ACO				✓	✓					
Sun et al [14]	ACO	✓		✓							
Liu et al [15]	OEMACS	✓	✓	✓							
Qin et al [16]	ACS-BVMP	✓					✓				
Alharbi et al [17]	ACS	✓									
Wang et al [18]	PSO	✓									
Benita Jacinth Suseela [19]	ACO-PSO	✓		✓				✓			
kang DONG [20]	ACO				✓				✓		
Dong et al [21]	DPGA				✓					✓	
Wu et al [22]	GA	✓									
Joseph et al [24]	GA	✓									✓
Torre et al [25]	NSGA-II	✓		✓							✓

algorithm to allocate PMs to VMs. This algorithm enhances the energy consumption and reduces SLAv, the number of active PMs, and the number of VM migrations.

In [44], the authors propose an algorithm based on FF and MBFD greedy algorithms to optimize VM-PM mapping. The authors propose a greedy approach for VM allocation that can maximize the energy efficiency of the cloud. The approach attempts to maximize VM consolidations on each PM to achieve greater energy efficiency than previous methods.

In addition, in [45], a greedy method is offered for placement of VMs with common memory pages on a PM. Program similarity across VMs is also considered as criteria to place them on a common PM.

In [46], the authors propose an evolutionary algorithm named Grey Wolf Optimization (GWO) for VM Placement (VMP) phase of VMC. This approach reduces the number of active PMs, energy consumption, SLAv, the number of migrations, and leads to the more efficient use of CPU and RAM resources.

In [47], the authors present an algorithm as an ILP problem for reducing energy consumption along with optimizing SLAv and performance. A similar method is proposed in [48] where MILP algorithm is employed for reducing energy consumption, SLAv and the number of migrations with a more efficient use of CPU resources. As shown in Table 2, significant research is conducted to address these metrics and offer various solutions.

TABLE 2. Summary of VM consolidation techniques.

Ref	Method	Energy/power	VM migrations	QOS / SLA	Resource	Resource wastage	Cost	Sleeping PMs
Jiang et al [36]	ABC	✓						
Mazumdar et al [39]	MILP		✓	✓			✓	
Zheng et al [40]	VMPMBBO	✓				✓		
Khan [41]	NVMC	✓	✓	✓	✓			
Aryania et al [42]	ACOS	✓	✓	✓				✓
Liu et al [49]	UACS	✓	✓	✓				

III. PROPOSED METHOD

In this paper, a multi-objective optimization based on ant colony is proposed for energy- and SLA-aware VM placement and consolidation. The proposed method delivers higher efficiency and improves the Quality of Service. The placement algorithm seeks optimal placement solutions by which the total consumed energy, resource wastage as well as the energy consumed by the traffic load of data exchange between VMs of a data center are minimized. In addition, the VM consolidation algorithm attempts to optimize resource usage through VM displacements in a data center while

TABLE 3. Definition of notations for proposed VM placement.

Notation	Definition
P	Power consumption
$P_{dynamic}$	Dynamic Power consumption
P_{static}	Static Power consumption
V	Voltage
F	Frequency
N	The number of virtual machine
R	The number of resources needed by each virtual machine
M	The number of physical machine
y_i	Denote PM_i is ative or inactive
Ut_j^p	efficiency of processor
$R_{i,1}^{VM}$	set of processors required by VM_i
x_{ij}	VM_i has been assigned or not to PM_j
(v_j, f_j)	The voltage and frequency pairs of processor
v_{highj}	The highest voltage of processor jth
f_{highj}	The highest frequency of processor jth
w	processor waste of each PM
R^{PM}	set of resources on PM_j
E	Energy consumption
$T_{i,1}^{NN}$	communication load matrix
$\tau_{v,p}$	value of pheromone
$n_{v,p}$	exploration element
C_p	capacity of each physical machine
b_p	load of each physical machine
r_v	number of requests in MIPS for VMs

focusing on reducing the energy consumption of the data center, lowering the VM displacements, minimizing SLA violations, and minimizing the number of active physical machines and achieving highest availability for users of these resources.

A. OBJECTIVE FUNCTION FOR PROPOSED VM PLACEMENT

In this section, standard definitions are presented for different concepts employed in our proposed method. These include the system architecture, task model, resources, and the multi-objective model for energy-aware task scheduling in green could data centers. Symbols used in these models proposed VM placement are listed in Table 3. In addition, it is assumed that the VMs, PMs and data center resources are heterogeneous and VMs are independent of each other. Cloud data centers host a set of computational resources on physical machines, which have been shared among a group of virtual machines. In the proposed method, VM placement task is considered as a multi-dimensional scheduling problem aimed at minimizing energy consumption, resource wastage and communication energy within a data center.

Objective 1: Minimizing power/energy consumption

Modern data centers contain a multitude of processors manufactured by CMOS technology. The power consumption in these devices can be categorized as dynamic and static power consumption, as expressed in Eq.

$$P = P_{dynamic} + P_{static} \tag{1}$$

Since the total power consumed to execute tasks also includes computation power by processors, the static part of power consumption can be ignored. The dynamic power consumption of processors can be calculated through Eq. 2 [6]:

$$P_{dynamic} = ACv^2f \tag{2}$$

where A is the percentage of active logic gates, C is the effective load capacitance, v is the supply voltage and f is the frequency of processor.

Nowadays, DVFS-enabled processors are employed to partly relieve the intensity of energy consumption in HPC systems [6]. DVFS-enabled processors can execute tasks using a discrete set of voltage and frequency pairs (v_j, f_j). Assume that each processor has k DVFS levels, that is, k processing operating points. Hence, supply voltage and frequency of processor j can be described by Eq. 3 where (v_{kj}, f_{kj}) is the voltage and frequency of processor j at level k.

$$(v_j, f_j) = \left\{ \begin{matrix} (v_{lowj}, f_{lowj}) = (v_{1j}, f_{1j}) < (v_{2j}, f_{2j}) < \\ \ldots < (v_{kj}, f_{kj}) = (v_{highj}, f_{highj}) \end{matrix} \right\} \tag{3}$$

In modern DVFS-equipped processors, the maximum power consumption of processor $P_{proc.highest}$ happens when it operates at maximum voltage $v_{highest}$ and frequency $f_{highest}$. The active power consumption of a processor under the voltage and frequency set (v_j, f_j) is calculated through Eq. 4.

$$P_{proc\,j} = P_{proc.highest} \times \frac{v_j^2 \times f_j}{v_{highest}^2 \times f_{highest}}$$
$$P_{proc.highest} = ACv_{highest}^2 f_{highest} \tag{4}$$

Since the task scheduling is performed for n DVFS-enabled processors, the total energy consumption can be calculated by Eq. 5.

$$P_{procs.active} = \sum\nolimits_{i=1}^{n} P_{proc.highest} (\sum\nolimits_{j=1}^{k} \frac{v_j^2 \times f_j}{v_{highest}^2 \times f_{highest}})$$
$$E_{procs.active} = \sum\nolimits_{i=1}^{n} P_{proc.highest} (\sum\nolimits_{j=1}^{k} \frac{v_j^2 \times f_j}{v_{highest}^2 \times f_{highest}} \times et\left(t_i, p_m(v_j, f_j)\right)) \tag{5}$$

In this model, N is the number of VMs, M is the number of PMs while R represents the set of resources (processors) required by each VM. The indicating variable y_j denotes wheter PM_j is active or inactive while X_{ij} indicates whether or not VM_i has been assigned to PM_j. The main goal is to minimize the data center's energy consumption. For this objective, the relation in [50] is expanded as Eq. 6:

$$P_j = (P_j^{busy} - P_j^{idle}) \times Ut_j^p + P_j^{idle} \tag{6}$$

where Ut_j^p is the efficiency of processor $Ut_j^p \in [0, 1]$ while P_j^{busy} and P_j^{idle} are the mean power values consumed when the j-th PM is active and idle, respectively. The overall consumed energy is given by Eq. 7 where $R_{i,1}^{VM}$ is the set of processors required by VM_i.

$$F_1 = Min \sum_{M}^{j=1} P_j^{PM} = \sum_{M}^{j=1} \Big[y_j \cdot \Big[(P_j^{busy} - P_j^{idle}) \times \sum_{N}^{i=1} (x_{ij} \cdot R_{i,1}^{VM}) + P_j^{idle}\Big]\Big] \tag{7}$$

Objective 2: Minimizing resource wastage

The other goal is to minimize resource waste in the data center. For this purpose, the relation proposed in [38] is expanded to define the related objective function. The processors are assumed to be wasted if they are not used by any virtual machine. Minimizing this waste of resources is the second objective of VM placement as expressed in Eq. 8. Here, W is the processor waste of each PM, R^{PM} is a set of resources on PM_j, and $R_{i,1}^{VM}$ indicates the set of processors required by VM_i.

$$F_1 = Min \sum_{M}^{j=1} P_j^{PM} = \sum_{M}^{j=1} \Big[y_j \cdot \Big[(P_j^{busy} - P_j^{idle}) \times \sum_{N}^{i=1} (x_{ij} \cdot R_{i,1}^{VM}) + P_j^{idle}\Big]\Big] \tag{8}$$

Objective 3: Minimizing energy cost of communication

Minimizing the energy cost of communication is the third objective of VM placement optimization, modeled as Eq. 9. A hierarchical topology is assumed between data center resources and the "k shortest paths" algorithm is employed to determine the network elements between two VMs. In Eq. 9, P_s^{busy} and P_s^{idle} are the energy consumption of the s-th network element when it is busy and idle, respectively. Also, $T_{i,1}^{NN}$ is the communication load matrix between VM_I and VM_1.

$$F_3 = Min \sum_{S}^{s=1} P_s^{NE} = \sum_{S}^{s=1} \Big[z_s \cdot \Big[\left(P_s^{busy} - P_s^{idle}\right) \times \sum_{N}^{i=1} \left(x_{is} \cdot T_{i,1}^{NN}\right) + P_s^{idle}\Big]\Big] \tag{9}$$

Further, the following constraints are taken into account in the energy-aware multi-objective optimization:

Constraint 1: $\sum_{M}^{j=1} x_{ij} = 1$ meaning each VM can only be hosted on one PM.

Constraint 2: $\sum_{N}^{i=1} R_{i,1}^{VM} \cdot x_{ij} \leq R_{j,1}^{PM} \cdot y_j$ This indicates that the number of resources allocated to VMs are less than or equal to the number of resources allocated to physical machines hosting VMs.

Constraint 3: $z_s, y_j, x_{ij} \in \{0, 1\}$

1) ANT-COLONY-BASED MULTI-OBJECTIVE VM PLACEMENT

VM placement optimization in a cloud data center is a multi-objective problem with several conflicting goals. These conflicting objectives may lead to a variety of solutions. Within the set of optimal solutions, no single solution has a greater overall performance than others. Multi-objective ant colony algorithm is one of the most widely used heuristic optimization methods which explore the feasible search space for optimal Pareto solutions of an optimization problem with conflicting goals. The AC algorithm employs principles such as elite selection and diversity maintenance across generations thus collecting the set of non-dominated solutions as the optimal Pareto solutions. A solution enters the Pareto front (or is called non-dominated) when the value of one objective can't be enhanced unless the value of another objective is deteriorated.

In a multi-objective optimization, solution $x^{(1)}$ dominates solution $x^{(2)}$ when both below conditions are met:

1. $x^{(1)}$ is not worse than $x^{(2)}$ for all objectives. Thus, the solutions are compared to each other according to the objective values (or based on the corresponding locations $z^{(1)}$ and $z^{(2)}$ in the target space).

2. $x^{(1)}$ is properly better than $x^{(2)}$ for at least one objective.

This definition applies for the two solution vectors. The dominance is however determined based on the objective vectors of the two solutions. All points not dominated by any other points are considered as Class 1 non-dominated points. A major property of the non-dominated solutions is that if a solution is superior to another for a given objective, the latter would be superior for at least one other objective. Therefore, none of these solutions can dominate the other thus being put to the same class. This characteristic leads to a diverse set of candidate points prior to selecting the final solution. Together, the mentioned collection of points is called as the non-dominated front.

The proposed multi-objective ant-colony-based placement algorithm obtains a Pareto front which contains the set of non-dominated solutions minimizing the overall multi-objective function. At each stage of the algorithm, a candidate is selected from a combination of the pheromones and newly-explored points.

The probability of physical machine p hosting the virtual machine v is given by Eq. 10. In this equation, $\tau_{v,p}$ is the value of pheromone in the set of virtual and physical machines.

$$P_P^V := \frac{\left[\tau_{v,p}\right]^{\alpha} \times \left[n_{v,p}\right]^{\beta}}{\sum \left[\tau_{v,p}\right]^{\alpha} \times \left[n_{v,p}\right]^{\beta}} \tag{10}$$

In Eq. 11, $n_{v,p}$ is the exploration element which yields solutions with the lowest resource wastage and energy consumption which is used in the decision formulae to obtain the solution. This exploring element is indeed the inverse of the difference between resource wastage and energy consumption in the sense that the VM-PM map with lower waste of resources and lower energy consumption has a higher $n_{v,p}$ value. Parameter C_p indicates the capacity and b_p denotes the load of each physical machine based on resource usage. Also, r_v is the number of requests in Million Instructions Per Second (MIPS) for VMs, E_j is the energy consumed by server j, and E_{max} is the maximum energy used by each server.

$$n_{v,p} := \frac{1}{\left|C_p - \left(b_p - r_v\right)\right|_1} + \frac{1}{\sum_{p}^{j=1} \frac{E_j}{E_{max}}} \tag{11}$$

The path defined by the pheromone series has a significant impact in the optimality of the solutions achieved by

TABLE 4. Parameters of the VM placement algorithm.

A	B	P	Ncycle	NAnts
1	2	0.5	10	5

ant colony method. To find a new optimal solution, the pheromone series is upgraded at each cycle. In Eq. 12, P is the pheromone evaporation parameter and $\tau_{v,p}$ is used to simulate the amount of evaporation to find the subset of solutions which mimize energy consumption and resource wastage.

$$\begin{aligned}\tau_{v,p} &:= (1-\rho)\times\tau_{v,p} + \frac{1}{f\,(S_{best})}, f\,(S_{best})\\ &= Min\sum\nolimits_{M}^{j=1} P_j^{PM} \times Min\sum\nolimits_{M}^{j=1} W_j^{PM}\\ &\quad\times Min\sum\nolimits_{S}^{s=1} P_s^{NE}\end{aligned} \tag{12}$$

Table 4 shows the parameters used in the simulations of the proposed placement algorithm. Parameters α and β are weighting factors assigned to pheromone and exploration elements, respectively. The parameter values are selected based on the performed simulations, the number of cycles and the number of ants as well as the relative significance of pheromone and exploration elements and the pheromone evaporation coefficient.

B. OBJECTIVE FUNCTION FOR PROPOSED VM CONSOLIDATION

Dynamic VM consolidation using dispersed VM grouping into minimum number of physical machines and shutting down the idle PMs within the cloud data centers will lead to a more efficient energy consumption. The main question in the VM consolidation problem pertains to which VMs should migrate.

Once the virtual machines are initially allocated to physical machines using the proposed VM placement method, the consolidation algorithm is run to dynamically group VMs into the least number of physical servers and thus achieve the optimization objectives. While the placement algorithm performs the initial assignment of VMs to PMs, the consolidation algorithm attempts, through multi-objective ant colony algorithm, to consolidate VMs to accomplish: 1) lowest energy consumption by the data center PMs, 2) lowest number of VMs, 3) minimum SLA violations, and 4) least number of active PMs.

Symbols used in these models proposed VM Consolidation are listed in Table 5.

Objective 1: Reduction of PM energy consumption

The first goal is to minimize the energy consumption of physical machines in heterogeneous cloud data centers. Given that the processors in a cloud data centers are heterogeneous and also equipped with DVFS technology, the associated energy consumptions can be obtained as Eq. 13.

$$F_1 = Min\sum\nolimits_{M}^{j=1} P_j^{PM} = \sum\nolimits_{M}^{j=1} \Big[y_j \cdot \Big[(P_j^{busy} - P_j^{idle}) \times \sum\nolimits_{N}^{i=1} (x_{ij}\cdot R_{i,1}^{VM}) + P_j^{idle}\Big]\Big] \tag{13}$$

TABLE 5. Definition of notations for proposed VM Consolidation.

Notation	Definition
N	Number of VMs
M	Number of PMs
R	Matrix that describes the set of resource needed by VM_i $[R_i^{VM}]_{N\times K}$
K	Number of resources available in a VM
R^{PM}	Matrix that describes the set of resource available PM_j $[R_{j,k}^{PM}]_{M\times K}$
$X_{i,j}$	1 if VM_i is assigned to PM_j , 0 if VM_i is not assigned
y_j	1 if PM_j is active, 0 if PM_j is idle

Objective 2: Minimizing Service Level Agreement violations

In cloud system, service providers oblige users for service level agreement (SLA) to ensure utilization rate of resources. In this agreement, different service level indices including the minimum CPU, RAM and storage capacities as well as the bandwidth. The number of SLA violations is a main criteria to be evaluated for any VM placement and consolidation approach and is calculated as [28]:

$$\begin{aligned}SLAV &= SLATAH.PDM\\ SLATAH &= \frac{1}{N}+\sum\nolimits_{i=1}^{N}\frac{T_{s_i}}{T_{a_i}} AndPDM = \frac{1}{M}+\sum\nolimits_{j=1}^{M}\frac{C_{d_i}}{C_{r_i}}\end{aligned} \tag{14}$$

where SLAV denotes SLA violation, SLATAH represents SLA violation Time per Active Host, and PDM stand for Performance Degradation due to Migrations. The following equation can be used to calculate SLATAH and PDM.

where N is the number of physical machines, Tsi is the time during which processors of physical machine i are 100% utilized whereas Tai is the active time of physical machine i. In Eq. 14, M is the number of VMs, Cdj is the estimated performance degradation of VM j due to migration, and Crj is the total capacity requested by the j.

Objective 3: Minimizing migration instances

Another important criterion that should be considered while evaluating a VM Consolidation approach is the number of VM migrations. Higher number of VM migrations leads to higher network load and energy consumption resulting in performance degradation. Eq. 15 can be used to calculate the number of migrations during a given time interval [28].

$$Migrations\,(F, t_1, t_2) = \sum\nolimits_{x=1}^{s}\int_{t_1}^{t_2} mig_x(F) \tag{15}$$

where F represents the current placements of VMs, $Mig_x(F)$ shows the number of migrations of server S_X within time interval $t_1 - t_2$ for the placement F.

Objective 4: Minimizing active physical machines

An additional goal is to lower energy consumption through minimizing the number of active physical machines.

TABLE 6. Parameters of ant-colony-based VM consolidation algorithm.

α	B	P	Ncycle	nAnts
0.1	0.9	0.1	2	5

1) MULTI-OBJECTIVE ANT COLONY OPTIMIZATION ALGORITHM

VM consolidation based on multi-objective ant colony optimization is based on Pareto method in which a set of non-dominated solutions minimizing the four objective functions are obtained. At each stage of the algorithm, a candidate solution is nominated as a cross between pheromone and exploration elements. The probabilistic decision making rule is based on Eq. 11 and the exploration element is calculated using Eq. 16.

$$n_{v,p} := \frac{1}{\left|C_p - \left(b_p - r_v\right)\right|_1 \times SLAV} \qquad (16)$$

In Eq. 16, C_p is the capacity of each physical machine and B_p indicates its processor utilization. Also, r_v is the MIPS requested for VM and SLAV indicates the number of SLA violations. This consists of the number of times the physical machine reaches 100% utilization of the processor.

The parameters used in the simulation of the proposed ant colony optimization algorithm are given in Table 6. Parameters α and β are weighting factors assigned to pheromone and exploration elements, respectively. The parameter values including the iteration number, ant population and weighting factors have been selected based on several performed experiments.

Fig.1 demonstrates the flowchart of the proposed algorithm for the multi-objective ant-colony-based placement and consolidation algorithm.

Given that the proposed algorithm consists of two phases of consolidation and placement, and in the proposed algorithm for each virtual machine these two phases must be performed, given that n is the number of virtual machines and m is the number of physical machines and for each virtual machine And for each k belonging to ant, the location of the virtual machines that can be hosted is checked, so the number of iterations is $(|n|\ (|k|\ |n|))$, and therefore the complexity of the algorithm is equal to $O(|n^2k|)$.

C. PERFORMANCE ANALYSIS WITH SIMULATION

Cloudism®is an open-source and accessible tool for modeling and simulation of cloud computing and distributed environments with resource provision capability. This tool supports the system and behavior of cloud elements including data centers, virtual machines and resource provision policies. In addition, it provides possibility to implement VM assignment techniques and policies in different could computing scenarios. The performance of the proposed MRAT-MACO algorithm is compared against those of single-objective virtual machines, i.e. First Fit Decreasing, dynamic voltage and frequency scaling and local regression, as well

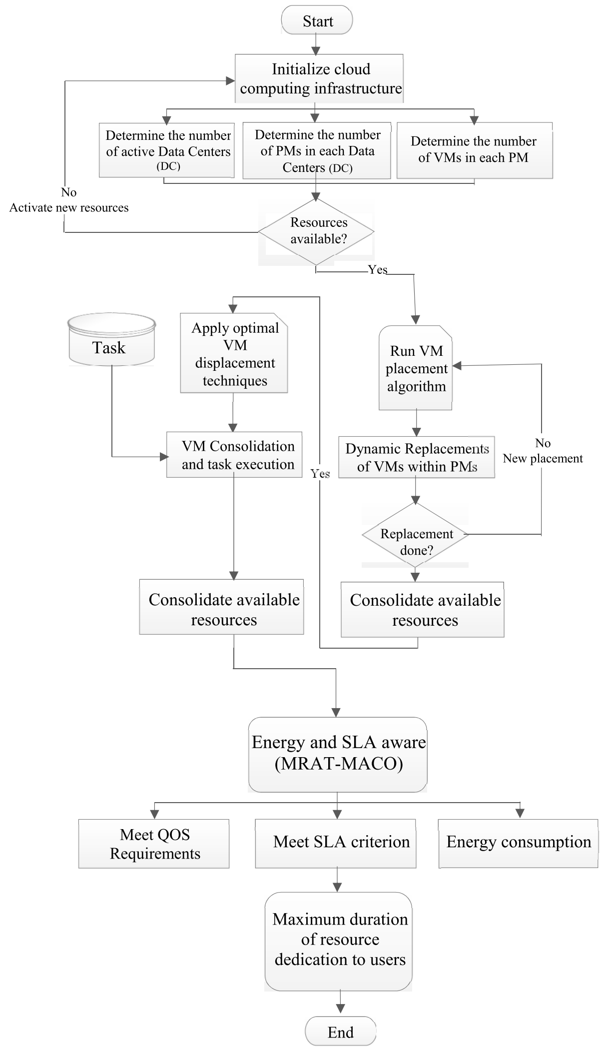

FIGURE 1. Flowchart of the proposed algorithm.

as multi-objective algorithms including multi-objective ant colony optimization and modified genetic algorithm.

For this purpose, the associated performance metrics include energy consumption in kW of power consumed by physical machines, resource wastage of processors on physical machines not used by any VM, number of active physical machines, communication energy cost resulting from the traffic load between VMs, and finally the execution time (in msec) of each algorithm.

D. PERFORMANCE ANALYSIS RESULTS FOR DIFFERENT VM PLACEMENT ALGORITHMS

For a test scenario with 700 hosts, the FFD algorithm uses all 700 physical machines with 14000 kW power consumption. For the same scenario, the multi-objective ant colony algorithm uses only 238 PMs with 9000kW power. In addition, the proposed MRAT-MACO approach yields the use of 700 PMs with 8700 kW power. The properties of the simulation environment are shown in Table 7.

TABLE 7. Properties of simulation environment.

Test specifications	specifications / value
Parameter settings	Setup
Simulator Datacenter Characteristics	MIPS: 1000 Storage: 1000000 RAM: 16 GB BW: 100 GB OS: Linux System Architecture: X86 Hypervisor: Xen
VMs properties	MIPS: 1000 Storage: RAM: 512 BW: 1000 Image Size: 3000 MB Hypervisor: Xen
Task Properties	Length: between (100 to 1000) PEs (Processing elements) number: 1 File size: 300 Output size: 300

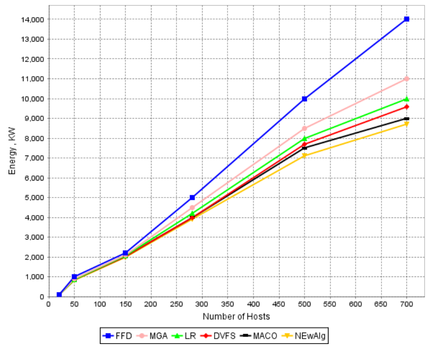

FIGURE 2. Comparison of energy consumption between VM placement algorithms.

Given the simulation results, it is seen that FFD and modified GA have the worst performance in terms of resource utilization while the multi-objective ant colony, local regression, Dynamic Voltage Frequency Scaling and MRAT-MACO deliver the best performances. In another test case with 20 PMs, the resource wastage for FFD, DVFS, LR and multi-objective ant colony, modified GA and MRAT-MACO is 40%, 5%, 3%, 4% and 4%, respectively. This means that in the worst case, 40% of the PM capacity is utilized. Further, the multi-objective ant colony algorithm achieves the best performance using all the capacity of PMs for mapping. For this case, 15 PMs are completely utilized for mapping. The results of performance analysis for different VM placement algorithms are shown in Table 8.

E. PERFORMANCE ANALYSIS OF TRI-OBJECTIVE VM CONSOLIDATION OPTIMIZATION

For the purpose of evaluating the proposed consolidation algorithm based on multi-objective ant colony optimization, the method is adjusted to be applied on a set of 700 PMs, 1000 VMs and 1500 tasks. The MRAT-MACO and multi-objective ant colony methods are compared against two single-objective VM consolidation methods, namely FFD and single threshold as well as two multi-objective methods, namely the modified GA and the multi-objective ant colony optimization proposed by Feller [10].

The performance metrics include power consumption of PMs in kW, number of displacements i.e. the number of PMs displaced for resource integration, number of SLA violation instances (that is the number of times a PM reaches 100% processor capacity), the percentage of SLA violations (that is the ratio of time where SLA violation persists to the total time of PM activity), and the algorithm execution time (msec). The first simulation set is executed focusing on three objectives, namely energy consumption, resource wastage and the communication energy cost. The simulation results for the test set are given in Table 9.

F. PERFORMANCE ANALYSIS OF QUAD-OBJECTIVE VM CONSOLIDATION OPTIMIZATION

The second simulation set is aimed at minimizing energy consumption, resource wastage, communication energy cost and the number of active physical machines. The simulation is carried out for up to 580 physical machines. The performance of the multi-objective ant colony and MRAT-MACO optimization are weighed against two single- Objective VM consolidation methods (FFD and single threshold) and two multi-objective algorithms (modified GA and the multi-objective ant colony optimization approach by Feller). The performance metrics are the energy consumption, number of VM displacements, number of SLA violations and the number of active PMs. The results of this set of simulations are shown in Table 10. In these simulations, the number of VMs is twice the number of hosts and each VM is dedicated to a given task.

G. PERFORMANCE ANALYSIS OF PROPOSED MULTI-OBJECTIVE ANT COLONY OPTIMIZATION ALGORITHM

After a thorough performance evaluation, it can be concluded that the proposed algorithm employing the multi-objective ant colony optimization does not always outperform other techniques.

In Fig. 2, the six examined algorithms are assessed in terms of the energy consumption metric. As the simulations indicate, the FFD method consumes the highest energy as it assigns each VM to the first PM with adequate resources without considering other possible PMs. The proposed MRAT-MACO algorithm, however, yields lower energy usage with its VM placement decisions.

Figures 3 and 4 illustrate the methods' comparison in terms of the number of active PMs and resource wastage. The results are indicative of a correlation between the number of active host PMs and the energy consumption. The higher The number of PMs used for hosting VMs, the greater the energy usage used for operating these devices.

TABLE 8. Performance analysis of different VM placement algorithms.

Hosts	Algorithm (policy)	Energy, KW	Resource Wastage (%)	# active PMs	Communication energy cost	Energy saving (%)
20	FFD	100	40	20	1500	0
	DVFS	88	5	16	0	23
	LR	89	3	17	100	38
	MACO	87	4	15	300	42
	MGA	90	20	19	500	40
	MRAT-MACO	85	4	14	500	43
50	FFD	1000	62	50	3000	0
	DVFS	890	4	38	1500	25
	LR	880	3	35	100	30
	MACO	860	3	35	1500	41
	MGA	900	30	49	2000	20
	MRAT-MACO	830	2	37	1700	43
150	FFD	2200	59	150	3500	0
	DVFS	2040	4	87	1800	30
	LR	2050	3	90	1000	31
	MACO	2020	3	78	1280	45
	MGA	2100	50	145	2500	25
	MRAT-MACO	2000	2	84	1240	45
280	FFD	5000	58	280	4000	0
	DVFS	4000	4	117	2000	30
	LR	4200	3	120	1300	31
	MACO	4000	3	108	1500	40
	MGA	4500	30	200	1500	20
	MRAT-MACO	3900	2	108	1000	41
500	FFD	10000	58	500	4100	0
	DVFS	7700	4	197	2500	30
	LR	8000	3	200	1800	31
	MACO	7500	3	188	2000	38
	MGA	8500	30	300	2000	20
	MRAT-MACO	7100	2	181	2400	40
700	FFD	1400	42	700	5500	0
	DVFS	9600	4	247	3000	30
	LR	10000	3	250	2800	31
	MACO	9000	3	238	3000	37
	MGA	11000	25	400	3000	25
	MRAT-MACO	8700	2	229	3300	40

Figures 5 and 6 demonstrate the connection between the communication energy and the percentage of the saved energy. The proposed approach has a better performance when the number of hosts is within 150 to 400 but it is outperformed by the multi-objective ant colony optimization for other cases. Also in terms of the percentage energy saved in VM placement, the proposed method achieves 42% saving while that of the modified GA is 25%.

The execution time of different VM placement algorithms for a change of PM hosts from 20 to 700 is shown in Fig. 7. The running time of the FFD algorithm remains nearly the same duration of 8000 msec from 20 hosts up to 700 hosts. Other placement techniques have longer running times with the runtime increasing linearly with the number of hosts. Since the multi-objective methods are more complicated that the single-objective methods, their execution time is shorter such that the dynamic voltage and frequency scaling as well as the local regression methods terminate at nearly the same duration which is much faster than the multi-objective ant colony approach. The running times for 700 hosts using MRAT-MACO, multi-objective ant colony, local regression and dynamic voltage and frequency scaling are 250, 41, 750, 39, 500, 39, 500 and 37 msec, respectively.

Figure 8 compares the results of the VM consolidation methods aimed to obtain the consumed energy. The FFD and single threshold methods are shown to have higher energy usage as compared to the modified GA, tri-objective ant colony approach, Feller multi-objective ant colony approach,

TABLE 9. Performance comparison between tri-objective VM consolidation algorithms.

Hosts	Algorithm (policy)	Energy, KW	# of migrations	# of SLA violations	% SLA violations
20	FFD	50	27	100	8.0
	ST	70	28	10	8.0
	MACO (Feller)	10	2	6	006.0
	MACO	10	4	7	007.0
	MGA	20	6	10	005.0
	MRAT-MACO	8	2	5	006.0
50	FFD	180	76	500	9.0
	ST	130	78	7	9.0
	MACO(Feller)	110	2	7	006.0
	MACO	160	6	7	007.0
	MGA	120	7	10	005.0
	MRAT-MACO	110	3	5	006.0
150	FFD	500	198	1000	9.0
	ST	450	196	10	9.0
	MACO(Feller)	380	2	9	006.0
	MACO	390	8	9	007.0
	MGA	400	8	28	005.0
	MRAT-MACO	370	7	8	006.0
280	FFD	1000	296	2000	9.0
	ST	1000	294	15	9.0
	MACO(Feller)	780	6	13	006.0
	MACO	790	10	7	007.0
	MGA	800	12	13	005.0
	MRAT-MACO	70	9	6	006.0
500	FFD	1650	799	4000	9.0
	ST	1600	797	8	9.0
	MACO(Feller)	1490	4	20	006.0
	MACO	1480	8	7	007.0
	MGA	1500	12	10	005.0
	MRAT-MACO	1400	10	5	
700	FFD	2500	999	6000	9.0
	ST	2450	998	8	9.0
	MACO(Feller)	1980	6	12	006.0
	MACO	1990	10	7	007.0
	MGA	2000	14	10	005.0
	MRAT-MACO	1800	12	5	006.0

and MRAT-MACO technique. As seen, the multi-objective ant colony and MRAT-MACO techniques yield the lowest energy consumption as they don't consider VM placement among physical machines whereas the meta-heuristic methods also take this into account.

In fig. 9, the algorithms are evaluated based on the number of VM displacements. As shown, the FFD method leads to higher VM displacements to achieve VM consolidation. In essence, when the FFD method receives a new VM displacement request, it attempts to find the first physical machine with sufficient resources to assign to the VM. If it does not find any active PM for VM allocation, it would activate a new PM without further considering the currently active VMs and PMs.

Fig. 10 draws a comparison between MRAT-MACO and other algorithms in terms of the number of displacements and migrations. The Feller multi-objective ant colony performs close to MRAT-MACO in achieving the lowest migrations. As indicated by the results, for a host count of 50 to 350, the MRAT-MACO algorithm results in lower migrations than multi-objective ant colony whereas for 350 to 700 hosts, the multi-objective ant colony leads to lower migrations compared to the proposed algorithm.

Figure 11 shows a comparison between the algorithms in terms of the number of SLA violations. As shown, the FFD method results in higher number of SLA violations as it attempts to assign as higher number of VMs as possible to each PM. Thus, the processor utilization may reach

TABLE 10. Performance comparison between quad-objective VM consolidation algorithms.

Hosts	Algorithm (policy)	Energy, KW	# of migrations	# of SLA violations	# of active PMs
20	FFD	10	10	100	1
	ST	10	10	100	1
	MACO (Feller)	9	5	5	1
	MACO (3 obj)	7	4	5	1
	MACO (4 obj)	5	5	5	1
	MGA	5	5	5	1
	MRAT-MACO	5	4	25	1
50	FFD	100	50	1500	20
	ST	900	200	2000	50
	MACO (Feller)	70	4	10	18
	MACO (3 obj)	80	5	20	16
	MACO (4 obj)	75	7	30	14
	MGA	75	10	50	12
	MRAT-MACO	70	3	100	50
150	FFD	300	100	2000	30
	ST	1500	250	4000	130
	MACO (Feller)	280	5	10	28
	MACO (3 obj)	290	4	20	26
	MACO (4 obj)	285	7	30	24
	MGA	285	10	50	42
	MRAT-MACO	283	4	100	60
280	FFD	1000	220	4000	100
	ST	3000	600	8000	290
	MACO (Feller)	990	6	10	98
	MACO (3 obj)	980	7	20	96
	MACO (4 obj)	975	8	30	94
	MGA	975	10	50	92
	MRAT-MACO	974	5	100	110
600	FFD	1100	300	5000	130
	ST	5000	1000	12000	400
	MACO (Feller)	1090	7	10	128
	MACO (3 obj)	1080	6	20	126
	MACO (4 obj)	1075	9	30	124
	MGA	1075	10	50	122
	MRAT-MACO	1065	5	100	164

100 percent in some servers leading to higher number of overloaded PMs and increased SLA violations.

Figure 12 presents the recorded number of SLA violations for multi-objective ant colony, the modified GA and MRAT-MACO. As observed, the multi-objective ant colony, the modified GA and MRAT-MACO result in lower number of SLA violations compared to the Feller multi-objective ant colony approach. This is because the metric of SLA violations is among the objectives in ant colony and modified GA approach whereas it is overlooked in Feller's algorithm.

Figures 13 and 14 shows the percentage of SLA violations committed by each algorithm. As inferred from Fig. 13, the multi-objective ant colony, modified GA and MRAT-MACO reduce the SLA violations compared to the FFD and single threshold approach. The percentage of SLA violations with respect to the number of active hosts demonstrated in Fig. 14.

The energy consumption values for different consolidation algorithms are shown in Fig. 15. Among these methods, the single threshold leads to the highest energy usage as it selects VMs and PMs in a random manner. In addition, the multi-objective ant colony (tri- and quad-objective), the proposed MRAT-MACO technique, and the modified GA have a similar performance in terms of energy consumption.

A comparison of the algorithm performances regarding VM migrations is shown in Fig. 16. Given their static, single-objective and single-solution nature, the single threshold and

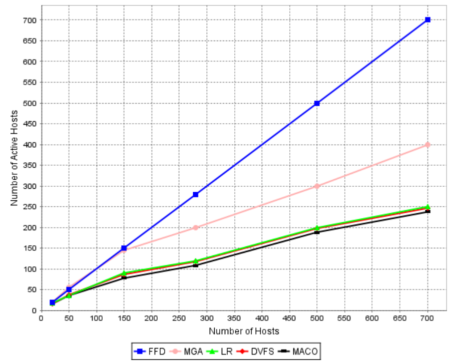

FIGURE 3. Comparison of number of active hosts between VM placement algorithms.

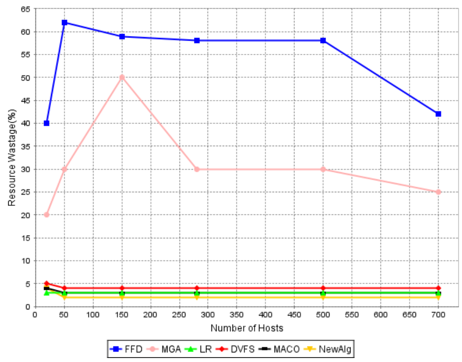

FIGURE 4. Comparison of percentage of resource wastage between VM placement algorithms.

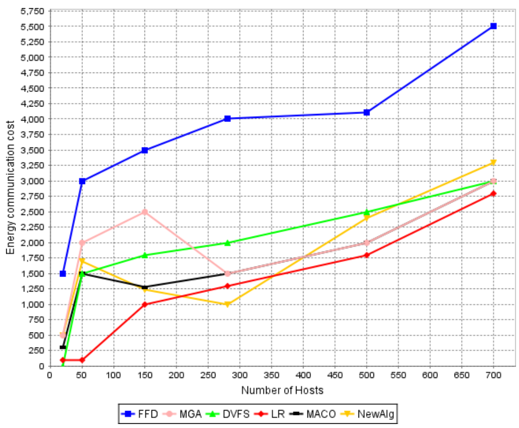

FIGURE 5. Comparison cost of communication energy between VM placement algorithms.

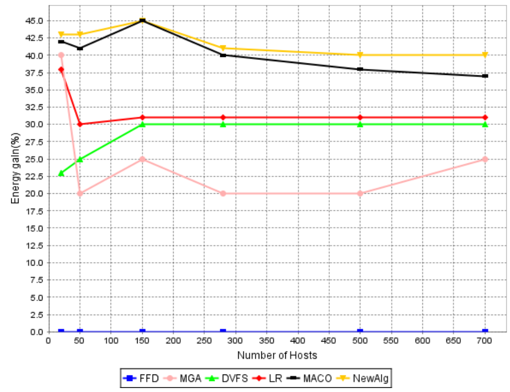

FIGURE 6. Comparison percentage of energy savings between VM placement algorithms.

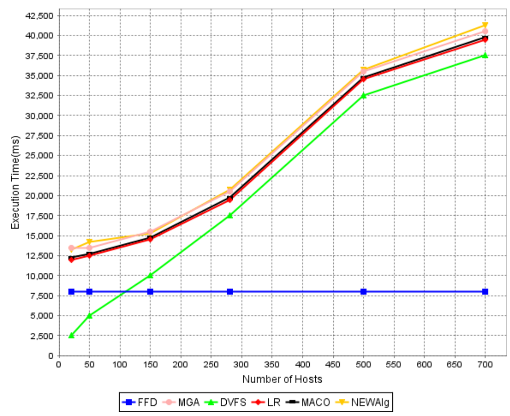

FIGURE 7. Comparison execution time between VM placement algorithms.

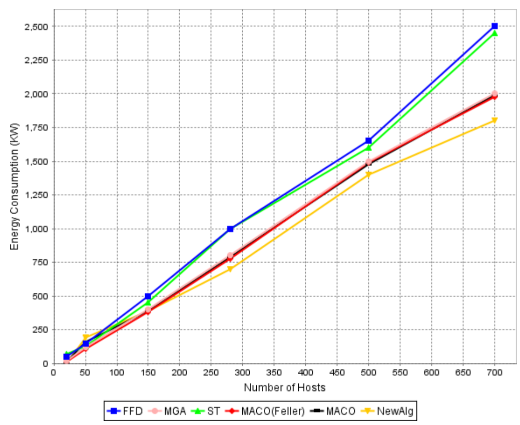

FIGURE 8. Comparison of energy consumption between VM consolidation algorithms.

FFD methods unsurprisingly lead to higher migrations compared to the multi-objective ant colony, modified GA and the proposed MRAT-MACO approach.

As seen in Fig. 17, the modified GA has a similar performance to the tri-objective ant colony algorithm with the least possible migrations in cases with limited number of hosts. However, as the number of hosts increases beyond, say 150, the GA performance deteriorates in terms of the migrations. For significant host numbers, the Feller's multi-objective ant colony and the proposed MRAT-MACO and the quad-objective ant colony approach demonstrate similar performances with the lowest number of migrations. In this case, the tri-objective ant colony optimization results performs better than the quad-objective approach because of its fewer achievable objectives.

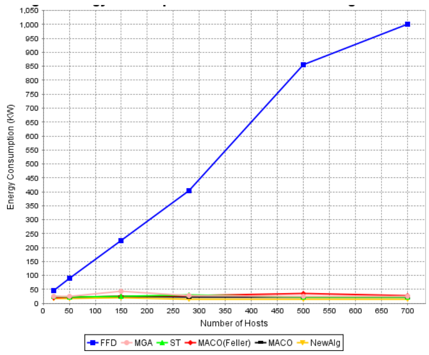

FIGURE 9. Comparison energy consumption between VM consolidation algorithms.

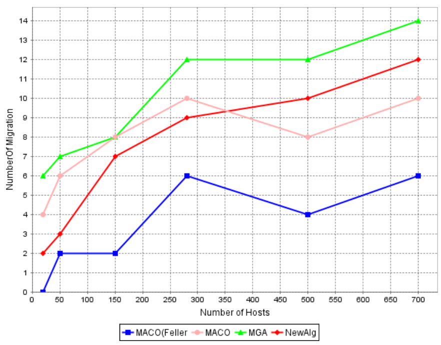

FIGURE 10. Comparison number of migrations in different VM consolidation algorithms.

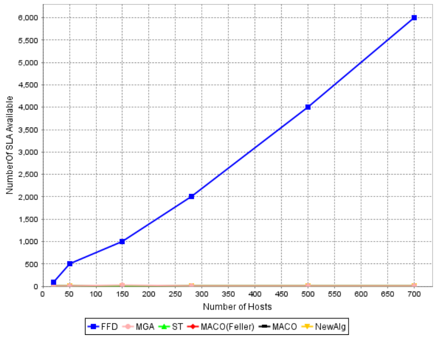

FIGURE 11. Comparison number of SLA violation between VM consolidation algorithms.

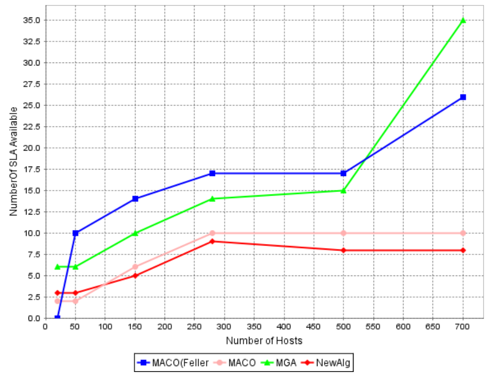

FIGURE 12. Comparison number of SLA violation.

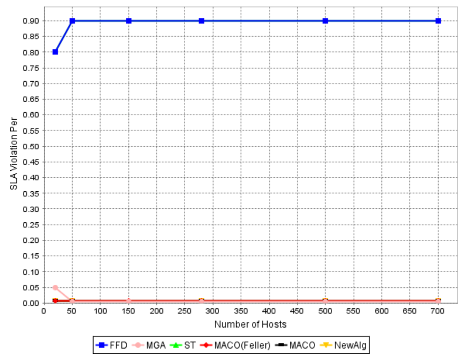

FIGURE 13. Comparison percentage of SLA violation between VM consolidation algorithms.

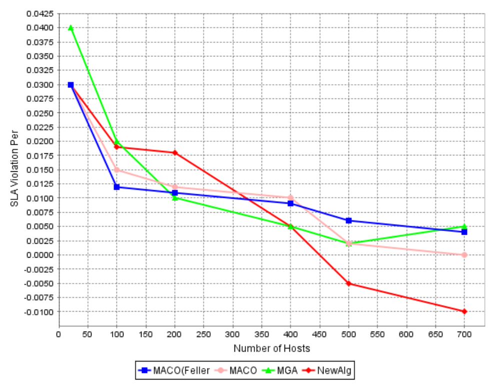

FIGURE 14. Comparison percentage of SLA violation.

In addition, Fig. 18 compares the consolidation algorithms in terms of SLA violation. The results show that the single-objective algorithms, namely the single threshold and FFD methods, lead to higher SLA violations as they don't consider SLA in their objective functions.

Figure 19 demonstrates that for higher number of hosts, the tri-objective ant colony achieves solutions with lowest SLA violation values. In addition, for a host number of 600, the quad-objective ant colony results in lower SLA violations than the modified GA approach. According to the obtained results, the Feller multi-objective approach as well as the proposed multi-objective ant colony and MRAT-MACO commit the highest SLA violations relative to the other three heuristic methods. The Feller multi-objective ant colony does not even consider SLA as an optimization objective.

In Fig. 20, the hosts used by the consolidation algorithms are shown. Among these methods, the modified GA achieves solutions with the lowest number of hosts. The results also indicate that the tri- and quad-objective ant colony methods yield lower number of hosts compared to the multi-objective Feller ant colony, MRAT-MACO and FFD approaches. The single threshold method requires higher number of hosts given its pre-dictated utilization rates. The static threshold

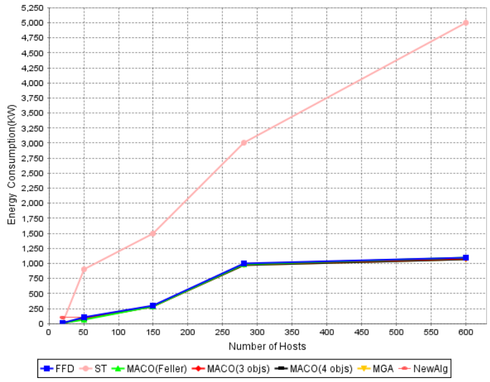

FIGURE 15. Comparison energy consumption between VM consolidation algorithms (tri- and quad-objective).

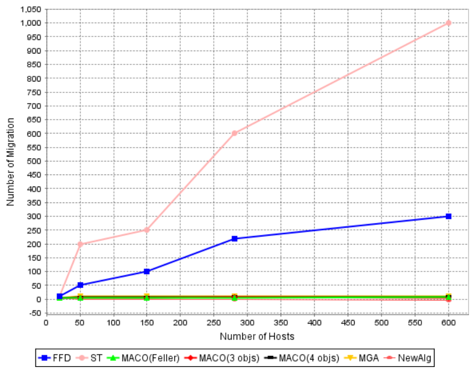

FIGURE 16. Comparison number of migrations between VM consolidation algorithms (tri- and quad-objective).

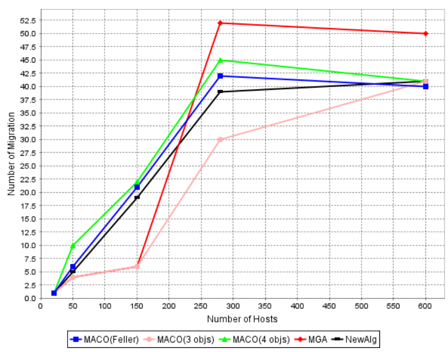

FIGURE 17. Comparison number of migrations between VM integrated consolidation algorithms (tri- and quad-objective).

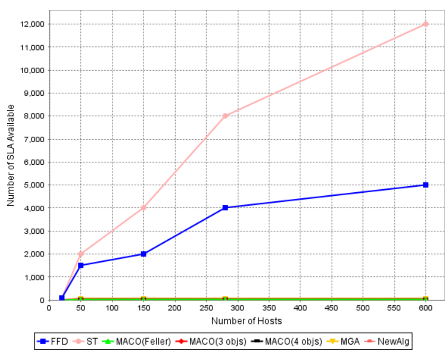

FIGURE 18. Comparison number of SLA violation between VM consolidation algorithms (tri- and quad-objective).

results in significant waste of resources consequently leading to further PM activations to meet VM demand.

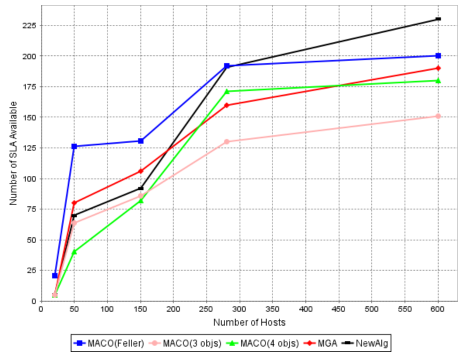

FIGURE 19. Comparison number of SLA violation between VM consolidation algorithms (tri- and quad-objective).

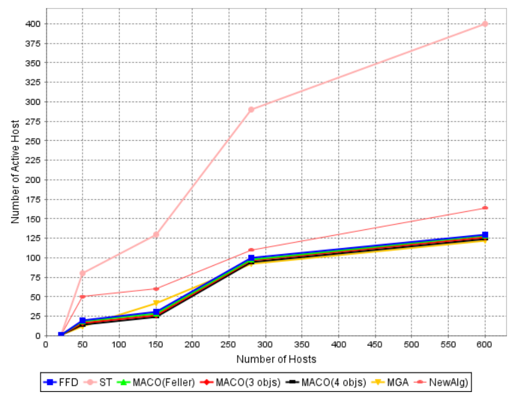

FIGURE 20. Comparison number of active host between VM consolidation algorithms (tri- and quad-objective).

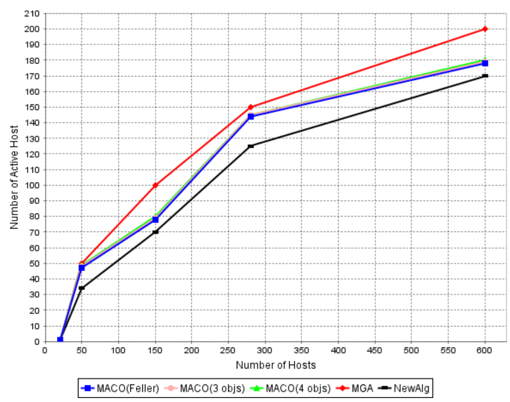

FIGURE 21. Comparison number of active host between VM consolidation algorithms (tri- and quad-objective).

Figure 21 compares the performance of multi-objective ant colony, modified GA and MRAT-MACO based on the number of activated hosts. To get a figure of the active hosts, simply the numbers of PMs (hosts) used by each algorithm are counted. For instance, in a test scenario with 50 physical machines in a data center, MRAT-MACO gives a consolidation solution with 34 active hosts while the Feller multi-objective approach uses 47, the tri-objective ant colony utilizes 48, the quad-objective ant colony employs 49 hosts, and the modified GA activates all 50 hosts.

Figure 22 demonstrates the running duration of each consolidation algorithm. As expected, the MRAT-MACO method has gained better running time compared to the

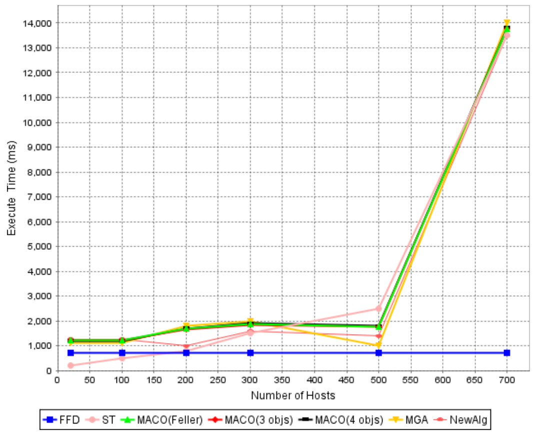

FIGURE 22. Comparison execution time between VM consolidation algorithms.

tri- and quad-objective ant colony algorithms. In addition, it is observed that the duration curves experience linear growth with the number of hosts.

IV. CONCLUSION

The proposed VM consolidation approach was evaluated by four metrics against two single-objective VM consolidation methods. The results of evaluations indicate that the FFT approach has the weakest performance in VM placement task whereas the proposed MRAT-MACO outperforms the other five approaches. Regarding the pursued objectives by the consolidation methods, those considering multiple objectives of energy saving, number of VM displacements and SLA violations, perform better than single-objective methods such as ST and FFT for all performance metrics.

Among the tri-objective metaheuristic methods, the proposed MRAT-MACO approach has a superior performance compared to the other three methods, i.e. Feller multi-objective ant colony optimization, modified GA, and multi-objective ant colony optimization yielding lower energy consumption and least SLA violations. The single-objective method of Single threshold delivers the weakest performance in terms of all mentioned criteria. Considering energy saving, the best-performing methods are the multi-objective ant colony optimization followed by the proposed MRAT-MACO and Feller's multi-objective ant colony methods. In terms of minimizing VM displacements and SLA violations, the multi-objective ant colony optimization and the proposed MRAT-MACO methods were the best-performing techniques.

In addition, the metaheuristic ant-colony optimization-based placement method performs better than the modified GA in terms of energy and SLA metrics whereas these techniques perform equally in terms of execution time for VM consolidation. Thus, increasing the number of objectives results in escalation of execution time and degraded performance for some of the considered metrics. This is attributed to the fact that these methods attempt to figure out solutions that fulfill all the requirements at the same time.

REFERENCES

[1] J. Masoudi, B. Barzegar, and H. Motameni, "Energy-aware virtual machine allocation in DVFS-enabled cloud data centers," *IEEE Access*, vol. 10, pp. 3617–3630, 2022, doi: 10.1109/ACCESS.2021.3136827.

[2] B. P. Rimal, E. Choi, and I. Lumb, "A taxonomy and survey of cloud computing systems," in *Proc. 5th Int. Joint Conf. INC, IMS IDC*, Aug. 2009, pp. 44–51.

[3] (2013). *World Energy Outlook*. Fact Sheet. [Online]. Available: http://goo.gl/Fxl.639

[4] R. Zhu, Z. Sun, and J. Hu, "Special section: Green computing," in *Future Generation Computer Systems*, vol. 28. New York, NY, USA: Elsevier, 2012, pp. 368–370.

[5] R. Zolfaghari and A. M. Rahmani, "Virtual machine consolidation in cloud computing systems: Challenges and future trends," *Wireless Pers. Commun.*, vol. 115, no. 3, pp. 2289–2326, Dec. 2020, doi: 10.1007/s11277-020-07682-8.

[6] B. Barzegar, H. Motameni, and A. Movaghar, "EATSDCD: A green energy-aware scheduling algorithm for parallel task-based application using clustering, duplication and DVFS technique in cloud datacenters," *J. Intell. Fuzzy Syst.*, vol. 36, pp. 5135–5152, Jan. 2019.

[7] E. Torre, J. J. Durillo, V. de Maio, P. Agrawal, S. Benedict, N. Saurabh, and R. Prodan, "A dynamic evolutionary multi-objective virtual machine placement heuristic for cloud data centers," *Inf. Softw. Technol.*, vol. 128, Dec. 2020, Art. no. 106390, doi: 10.1016/j.infsof.2020.106390.

[8] S. K. Mishra, D. Puthal, B. Sahoo, P. P. Jayaraman, S. Jun, A. Y. Zomaya, and R. Ranjan, "Energy-efficient VM-placement in cloud data center," *Sustain. Comput., Inform. Syst.*, vol. 20, pp. 48–55, Dec. 2018.

[9] M. R. Mesbahi, A. M. Rahmani, and M. Hosseinzadeh, "Reliability and high availability in cloud computing environments: A reference roadmap," *Hum.-centric Comput. Inf. Sci.*, vol. 8, no. 1, Dec. 2018, doi: 10.1186/s13673-018-0143-8.

[10] E. Feller, L. Rilling, and C. Morin, "Energy-aware ant colony based workload placement in clouds," in *Proc. IEEE/ACM 12th Int. Conf. Grid Comput.*, Sep. 2011, pp. 26–33.

[11] Y. Gao, H. Guan, Z. Qi, Y. Hou, and L. Liu, "A multi-objective ant colony system algorithm for virtual machine placement in cloud computing," *J. Comput. Syst. Sci.*, vol. 79, no. 8, pp. 1230–1242, Dec. 2013.

[12] V. A. K. Sarma, R. Rajendra, P. Dheepan, and K. S. Kumar, "An optimal ant colony algorithm for efficient VM placement," *Indian J. Sci. Technol.*, vol. 8, no. S2, pp. 156–159, 2015.

[13] T. P. Shabeera, S. D. M. Kumar, S. M. Salam, and K. M. Krishnan, "Optimizing VM allocation and data placement for data-intensive applications in cloud using ACO metaheuristic algorithm," *Eng. Sci. Technol. Int. J.*, vol. 20, no. 2, pp. 616–628, Apr. 2017.

[14] X. Sun, K. Zhang, M. Ma, and H. Su, "Multi-population ant colony algorithm for virtual machine deployment," *IEEE Access*, vol. 5, pp. 27014–27022, 2017.

[15] X.-F. Liu, Z.-H. Zhan, J. D. Deng, Y. Li, T. L. Gu, and J. Zhang, "An energy efficient ant colony system for virtual machine placement in cloud computing," *IEEE Trans. Evol. Comput.*, vol. 22, no. 1, pp. 113–128, Feb. 2018.

[16] Y. Qin, H. Wang, F. Zhu, and L. Zhai, "A multi-objective ant colony system algorithm for virtual machine placement in traffic intense data centers," *IEEE Access*, vol. 6, pp. 58912–58923, 2018.

[17] F. Alharbi, Y.-C. Tian, M. Tang, W.-Z. Zhang, C. Peng, and M. Fei, "An ant colony system for energy-efficient dynamic virtual machine placement in data centers," *Expert Syst. Appl.*, vol. 120, pp. 228–238, Apr. 2019.

[18] S. Wang, Z. Liu, Z. Zheng, Q. Sun, and F. Yang, "Particle swarm optimization for energy-aware virtual machine placement optimization in virtualized data centers," in *Proc. Int. Conf. Parallel Distrib. Syst.*, Dec. 2013, pp. 102–109.

[19] B. B. J. Suseela and V. Jeyakrishnan, "A multi-objective hybrid ACO-PSO optimization algorithm for virtual machine placement in cloud computing," *Int. J. Res. Eng. Technol.*, vol. 3, no. 4, pp. 474–476, Apr. 2014.

[20] J.-K. Dong, H.-B. Wang, Y.-Y. Li, and S.-D. Cheng, "Virtual machine placement optimizing to improve network performance in cloud data centers," *J. China Univ. Posts Telecommun.*, vol. 21, no. 3, pp. 62–70, 2014.

[21] X.-F. Liu, Z.-H. Zhan, and J. Zhang, "An energy aware unified ant colony system for dynamic virtual machine placement in cloud computing," *Energies*, vol. 10, no. 5, p. 609, May 2017.

[22] Y.-S. Dong, G.-C. Xu, and X.-D. Fu, "A distributed parallel genetic algorithm of placement strategy for virtual machines deployment on cloud platform," *Sci. World J.*, vol. 2014, pp. 1–12, Oct. 2014.

[23] G. Wu, M. Tang, Y. C. Tian, and W. Li, "Energy-efficient virtual machine placement in data centers by genetic algorithm," in *Proc. Int. Conf. Neural Inf. Process.* Berlin, Germany: Springer, Nov. 2012, pp. 315–323.

[24] C. T. Joseph, K. Chandrasekaran, and R. Cyriac, "A novel family genetic approach for virtual machine allocation," *Proc. Comput. Sci.*, vol. 46, pp. 558–565, Jan. 2015.

[25] E. Torre, J. J. Durillo, V. de Maio, P. Agrawal, S. Benedict, N. Saurabh, and R. Prodan, "A dynamic evolutionary multi-objective virtual machine placement heuristic for cloud data centers," *Inf. Softw. Technol.*, vol. 128, Dec. 2020, Art. no. 106390, doi: 10.1016/j.infsof.2020.106390.

[26] F. F. Moges and S. L. Abebe, "Energy-aware VM placement algorithms for the OpenStack neat consolidation framework," *J. Cloud Comput.*, vol. 8, no. 1, pp. 1–14, Dec. 2019.

[27] E. Parvizi and M. H. Rezvani, "Utilization-aware energy-efficient virtual machine placement in cloud networks using NSGA-III meta-heuristic approach," *Cluster Comput.*, vol. 23, no. 4, pp. 1–23, 2020.

[28] A. Beloglazov and R. Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud computings," *Concurrency Comput., Pract. Exper.*, vol. 24, no. 13, pp. 397–420, 2012.

[29] H. M. Naeen, E. Zeinali, and A. T. Haghighat, "A stochastic process-based server consolidation approach for dynamic workloads in cloud data centers," *J. Supercomput.*, vol. 76, no. 3, pp. 1903–1930, Mar. 2020.

[30] Y. Qiu, C. Jiang, Y. Wang, D. Ou, Y. Li, and J. Wan, "Energy aware virtual machine scheduling in data centers," *Energies*, vol. 12, no. 4, p. 646, Feb. 2019.

[31] L. Xie, S. Chen, W. Shen, and H. Miao, "A novel self-adaptive VM consolidation strategy using dynamic multi-thresholds in IaaS clouds," *Future Internet*, vol. 10, no. 6, pp. 1–18, 2018.

[32] A. Pahlavan, M. Momtazpour, and M. Goudarzi, "Power reduction in HPC data centers: A joint server placement and chassis consolidation approach," *J. Supercomput.*, vol. 70, no. 2, pp. 845–879, 2014.

[33] A. Roytman, A. Kansal, S. Govindan, J. Liu, and S. Nath, "PACMan: Performance-aware virtual machine consolidation," in *Proc. 10th Int. Conf. Autonomic Comput. (ICAC)*, 2013, pp. 83–94.

[34] A. Ullah, J. Li, Y. Shen, and A. Hussain, "A control theoretical view of cloud elasticity: Taxonomy, survey and challenges," *Cluster Comput.*, vol. 21, pp. 1735–1764, Dec. 2018.

[35] J. N. Witanto, H. Lim, and M. Atiquzzaman, "Adaptive selection of dynamic VM consolidation algorithm using neural network for cloud resource management," in *Future Generation Computer Systems*. New York, NY, USA: Elsevier, 2018, pp. 1–20.

[36] J. Jiang, Y. Feng, J. Zhao, and K. Li, "DataABC: A fast abc based energy-efficient live VM consolidation policy with data-intensive energy evaluation model," *Future Gener. Comput. Syst.*, vol. 74, pp. 132–141, Sep. 2017.

[37] M.-H. Malekloo, N. Kara, and M. El Barachi, "An energy efficient and SLA compliant approach for resource allocation and consolidation in cloud computing environments," *Sustain. Comput., Inform. Syst.*, vol. 17, pp. 9–24, Mar. 2018, doi: 10.1016/j.suscom.2018.02.001.

[38] H. Xiao, Z. Hu, and K. Li, "Multi-objective VM consolidation based on thresholds and ant colony system in cloud computing," *IEEE Access*, vol. 7, pp. 53441–53453, 2019, doi: 10.1109/ACCESS.2019.2912722.

[39] S. Mazumdar and M. Pranzo, "Power efficient server consolidation for cloud data center," *Future Gener. Comput. Syst.*, vol. 70, pp. 4–16, May 2017.

[40] Q. Zheng, R. Li, X. Li, N. Shah, J. Zhang, F. Tian, K.-M. Chao, and J. Li, "Virtual machine consolidated placement based on multi-objective biogeography-based optimization," *Future Gener. Comput. Syst.*, vol. 54, pp. 95–122, Jan. 2016.

[41] M. A. Khan, "An efficient energy-aware approach for dynamic VM consolidation on cloud platforms," *Cluster Comput.*, vol. 24, no. 4, pp. 3293–3310, Dec. 2021, doi: 10.1007/s10586-021-03341-0.

[42] A. Aryania, H. S. Aghdasi, and L. M. Khanli, "Energy-aware virtual machine consolidation algorithm based on ant colony system," *Grid Comput.*, vol. 16, no. 3, pp. 477–491, 2018.

[43] F. F. Moges and S. L. Abebe, "Energy-aware VM placement algorithms for the OpenStack neat consolidation framework," *J. Cloud Comput.*, vol. 8, no. 1, pp. 1–14, Dec. 2019.

[44] X. Zhang, T. Wu, M. Chen, T. Wei, J. Zhou, S. Hu, and R. Buyya, "Energy-aware virtual machine allocation for cloud with resource reservation," *J. Syst. Softw.*, vol. 147, pp. 147–161, Jan. 2019.

[45] A. Beloglazov, J. Abawajy, and R. Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," *Future Gener. Comput. Syst.*, vol. 28, no. 5, pp. 755–768, May 2012.

[46] A. Al-Moalmi, J. Luo, A. Salah, and K. Li, "Optimal virtual machine placement based on grey wolf optimization," *Electronics*, vol. 8, no. 3, p. 283, 2019.

[47] K. Chang, S. Park, H. Kong, and W. Kim, "Optimizing energy consumption for a performance-aware cloud data center in the public sector," *Sustain. Comput., Informat. Syst.*, vol. 20, pp. 34–45, Dec. 2018.

[48] R. Nasim, E. Zola, and A. J. Kassler, "Robust optimization for energy-efficient virtual machine consolidation in modern datacenters," *Cluster Comput.*, vol. 21, no. 3, pp. 1681–1709, Sep. 2018.

[49] X.-F. Liu, Z.-H. Zhan, and J. Zhang, "An energy aware unified ant colony system for dynamic virtual machine placement in cloud computing," *Energies*, vol. 10, no. 5, p. 609, May 2017.

[50] W. Leinberger, G. Karypis, and V. Kumar, "Multi-capacity bin packing algorithmswith applications to job scheduling under multiple constraints," in *Proc. Int. Conf. Parallel Process.*, Sep. 1999, pp. 404–412.

[51] L. Li, J. Dong, D. Zuo, and J. Wu, "SLA-aware and energy-efficient VM consolidation in cloud data centers using robust linear regression prediction model," *IEEE Access*, vol. 7, pp. 9490–9500, 2019, doi: 10.1109/ACCESS.2019.2891567.

[52] S. K. Roy, R. Devaraj, A. Sarkar, and D. Senapati, "SLAQA: Quality-level aware scheduling of task graphs on heterogeneous distributed systems," *ACM Trans. Embedded Comput. Syst.*, vol. 20, no. 5, pp. 1–31, Sep. 2021, doi: 10.1145/3462776.

BADIEH NIKZAD received the B.S., M.S., and Ph.D. degrees in computer engineering from Islamic Azad University, Babol, Iran, in 2014, 2017, and 2022, respectively. Her research interests include cloud computing and task scheduling.

BEHNAM BARZEGAR received the B.S. degree in computer engineering from Islamic Azad University, Sari, Iran, in 2006, the M.S. degree in computer engineering from Islamic Azad University, Najafabad, in 2009, and the Ph.D. degree in computer engineering from Islamic Azad University, in 2018. He is currently an Associate Professor with the Department of Computer Engineering, Islamic Azad University, Babol Branch, Babol, Iran. His research interests include green cloud computing, task scheduling, and formal methods (Petri net).

HOMAYUN MOTAMENI received the B.S. degree in computer engineering from Shahid Beheshti University, Tehran, Iran, in 1995, and the M.S. and Ph.D. degrees in computer engineering from the Science and Research Branch, Islamic Azad University, Tehran, in 1998 and 2007, respectively. He is currently an Associate Professor with the Department of Computer Engineering, Islamic Azad University, Sari Branch, Sari, Iran. His research interests include software engineering, performance analysis, evolutionary computation, fuzzy systems, and green cloud computing.