

An adaptive scheduling framework for the dynamic virtual machines placement to reduce energy consumption in cloud data centers

Abdulrahman Nahhas
 Faculty of computer science
 Otto von Guericke University
 Magdeburg, Germany
abdulrahman.nahhas@ovgu.de

Jahnavi Thimmaiah Cheyyanda
 Ableton AG
 Berlin, Germany
jahnavi.cheyyanda@gmail.com

Klaus Turowski
 Faculty of computer science
 Otto von Guericke University
 Magdeburg, Germany
klaus.turowski@ovgu.de

Abstract

Cloud computing has revolutionized the IT industry through its on-demand provisioning of virtualized resources through the internet. Although it relies on sharing of resources to improve the performance of datacenters, it has increased the complexity of IT systems in recent years. To meet the market requirements, cloud providers are expanding their datacenters with a large number of servers leading to high energy consumption and therefore, increasing the carbon footprint. Environmental impact and rapidly surging energy costs have become a major concern for both the government bodies and the IT service providers. In this paper, we propose a genetic algorithm based hybrid load management strategy which uses multiple existing virtual machines allocation policies to minimize the energy consumption, the violations in the service level agreement and the number of migrated virtual machines. The presented solution approach is evaluated on CloudSim Plus simulation framework using the well known PlanetLab workload. The results obtained from the experiments show substantial improvement in energy consumption in comparison to the individual approaches while maintaining the performance constraints.

Keywords: Adaptive load distribution strategies, energy-aware virtual machines allocation, heuristic and meta-heuristic optimization, dynamic consolidation.

1. Introduction

Cloud computing has emerged as one of the remarkable technologies in the recent years. Introduction of cloud computing technology in the IT industry has sparked a revolution and has resulted in a substantial change in the design and management of IT infrastructure. Cloud computing paradigm offers on-demand resources for computing, storage, networking and software over the internet [1]. The

pay-as-you-go payment model provided by many public cloud providers is an appealing feature for many cloud users. With this model, the cloud users have to pay only for the services they want to use. This allows companies to save a lot on their capital investment and management of IT infrastructure. According to the gartner hype cycle, cloud computing has already passed the peak of inflated expectations in 2012, and as of 2019 it is at the Slope of Enlightenment [2]. Cloud computing is been adapted by numerous business domains as it provides greater stability, helps in cost-cutting and offers enhanced scalability and security [3]. The question is no longer whether cloud computing is a suitable technology for the business but how it can be utilized to make maximum profit.

Virtualization is the technology that enables seamless management of resources irrespective of their physical location and allows cloud providers to supply resources in a cost-effective manner. While physical machines come with various computing capacities [4] users of cloud demand Virtual Machines (VM) with distinct processing, memory, storage or networking requirements. This problem is often referred to as virtual machine placement problem. Therefore, the mapping of resources should be done carefully for maximum resource utilization, fewer violations Service Level Agreement (SLA) and low energy consumption. Although cloud computing provides boundless advantages over traditional data centers, it has created a lot of challenges to the cloud providers and researchers. There are two main motivations for our research in the field of energy efficiency of cloud data centers. The first one is the operational cost. Since cloud computing operates on a huge pool of physical machines it demands an enormous amount of energy. Currently, data centers contribute to 3% of overall electricity consumption globally [5]. With the increasing energy cost, the operational costs are exponentially increasing in large IT-landscapes. The second reason is the increasing carbon footprint. The Information and Communication Technology (ICT)

sector is estimated to contribute to around 2% of the total greenhouse gas emissions. The authors in [5] believe that around 3 - 13% of global electricity in 2030 will be consumed by data centers. Due to enormous need for computing, storage and other services data centers will grow to become one of the most important contributors to the rising atmospheric CO₂ levels among other commercial electricity consumers [6]. Following the recent developments in climate change, governments are taking measures to mitigate the affects of CO₂ emissions. This puts organisations under a great pressure to find sustainable solutions to cut down their carbon footprint.

There exist many optimization techniques for VM allocation and the prominent ones are the heuristic techniques. Although these approaches have shown good results for VM allocation, they are specific to solving only one kind of optimization problem such as reducing the power consumption or improving the SLA or in the data center. Another commonly used approach is the meta-heuristic approach which is mostly inspired by nature and is found to produce high quality and robust solutions [7]. However, this approach requires high computational effort. In this paper, we propose a novel hybrid load management technique which uses both heuristic and meta-heuristic approaches to reduce the energy consumption of servers, SLA violations and number of VM migrations in a cloud-based data center through effective VM consolidation. Our approach is directed towards incorporating several existing algorithms and not creating a new algorithm for VM allocation. It is based on the observation that different algorithms are best suited for minimization of power and minimization of SLA violations. This approach will allow choosing of VM consolidation algorithms, according to the need of cloud users.

In this work, we aim to address the following research question.

RQ: Does a hybrid load management strategy using heuristic and meta-heuristic approaches perform better when compared to a specific or generic approach for reducing energy consumption in data centers?

The question is based on the hypothesis presented in [7] that an optimized combination of different heuristic algorithms used during a specified time span would outperform an individual algorithm for sophisticated management of IT resources in data centers.

The remainder of this paper is organized as follows. In section 2, we present the state-of-the-art that details the approaches for energy-aware load management in cloud-based data centers. We analyse the outcomes of these research works and our perspective on the solutions. The design and theoretical framework of our

solution in the form of a conceptual model is discussed in section 3. In the next section, we demonstrate the experiments conducted through simulation on Cloudsim Plus and evaluate the results. Finally, we conclude by discussing the outcomes of the experiments and outline the future work of this research.

2. State of the art

In the last two decades, researchers have made enormous contribution to improve the VM placement and consolidation problem to optimize the resource utilization and power consumption in the data centers, while maintaining the performance constraints. The optimization techniques can be broadly classified into four different 11 approaches. They are heuristics, meta-heuristics, deterministic algorithms and the hybrid approaches. Following are some of the approaches designed to address the VM placement and consolidation problem which are discussed in the state of the art publications to improve energy efficiency.

2.1. Heuristic approaches

VM allocation can be depicted to a form of bin packing problem where the virtual machines modelled as items are placed on a minimum number of physical machines modelled as bins [8]. Various resources of a physical machines such as the CPU, memory, network and storage of a physical machine are taken into consideration during the VM placement. Due to the complexity of the VM consolidation problem, it is considered to be NP-hard [9] [7] [10]. This led to the investigation of a many heuristic-based approaches to optimize the operations in the cloud datacenter.

Heuristic techniques are defined as, the optimization techniques which involves quick and practical methods to solve complex problems. However, these techniques do not guarantee optimal solutions but often yield satisfactory outcomes. A heuristic based, energy-aware resource allocation approach is introduced in [11]. The authors have proposed one of the finest frameworks by considering heterogeneous infrastructure with various levels of SLAs and is independent of workload types. The proposed Modified Best Fit Decreasing algorithm sorts all VMs in decreasing order of their current CPU utilization and allocates the VMs to the most power efficient nodes first. However, they have considered only CPU to consume most part of the energy in a physical machine. Similar to [11], the research work done in [12] aims to improve the energy efficiency of the data centers while improving the SLA violations. To assign a VM to a suitable physical server the authors suggest Best Fit Decreasing algorithm. When the CPU

utilization of a server exceeds a certain predetermined upper threshold, the VM with highest CPU utilization needs to be migrated to another server in order to reduce the total number of migrations. Overall, the proposed method reduces energy consumption, total execution time, and SLA violations and thereby reduces the operational costs involved. The work done in [13] tries to solve the problem of energy consumption during VM consolidation by reducing the number of VM migrations and improving the utilization of resources. The proposed algorithm considers identifying overloaded hosts, choosing the right VM to be migrated and selection of the best host on which the VM can be placed. The authors have compared the proposed approach against several power-aware VM management techniques that are available in Cloudsim simulation tool. The obtained results showed improvements in terms of energy consumption.

An application placement controller called pMapper is proposed in [14] to dynamically place applications on an optimal set of physical servers to minimize power consumption while meeting performance guarantees. Servers that are not being used for a longer period of time can be switched off or turned to low power states to reduce power consumption. Through their experiments they have established that pMapper is the best solution over many other Power aware and Power unaware algorithms for reducing energy consumption, number of VM migrations and for improving the Quality of Service (QoS). An advancement to the VM placement algorithm for OpenStack Neat framework was proposed in [15]. To improve the energy efficiency, SLA violations and number of VM migrations power efficient algorithms with bin-packing heuristic called Power Efficient First-Fit Decreasing (PEFFD) and Power Efficient Best-Fit Decreasing (PEBFD) are introduced considering only CPU to consume most of the energy. In addition, a new bin-packing heuristic called a Medium-Fit (MF) is developed to reduce SLA violation and VM migrations. Like most of the research work done so far they have taken only CPU utilization into consideration. They also mention that peak power of hosts must be known for practical implementation of the proposed algorithms.

2.2. Meta-heuristics approaches

Meta-heuristics are high level heuristics used to obtain a satisfactory solution with limited information and resources. Search algorithms which are classified under evolutionary algorithms are generally used in meta-heuristic optimization. Genetic Algorithm [16], Simulated Annealing (SA) [17], Ant Colony

Optimization (ACO) [18] and Particle Swarm Optimization (PSO) [19] are some of the commonly used algorithms for VM consolidation.

Genetic Algorithm is a popular meta-heuristic approach in which a population of possible candidate solutions known as individuals are evolved over many generations to reach an optimal solution for a known problem. The optimization starts with random individuals and undergoes several genetic operations to reach a near global optimum. For instance, in [20] the authors have presented a solution approach for load balancing using Genetic algorithms. The results show that Genetic algorithms performed the best when compared to Greedy and Round-Robin algorithm. It also reduced the number of VM migrations as the VMs were allocated to physical machines based on a fitness function. In [21], authors have studied various resource allocation methods in order to find the more efficient technique that utilizes the resources effectively and confine to the SLA which is agreed between the customer and the service provider. The study concludes that adaptive resource management scheme such as genetic algorithm provides better solution for resource allocation in cloud computing. In the referenced paper [10], the VM placement problem is considered to be a multi-objective optimization problem. The three objectives that are considered simultaneously are minimizing total resource wastage, power consumption and thermal dissipation costs. An improved grouping genetic algorithm is used for efficiently searching global optimal solutions. To evaluate the solutions obtained by the grouping genetic algorithm, a fuzzy-logic based evaluation approach is used for combining different objectives. The fuzzy rule considers a solution which results in small resource wastage, low power consumption and low temperature as a good solution”[10].

Simulated annealing is a global optimization technique that is based on the metal annealing processing [17]. Virtual machine placement problem was addressed in the master thesis done in [4] using a utilization based genetic algorithm (UBGA) and utilization based simulated annealing (UBSA). The genetic algorithm based approach uses a novel fitness function which considers resource wastage, network load and energy consumption. The chromosome structure matches the tree representation of datacenter network topology. The leaves represent physical machines in the datacenter and the list of virtual machines are denoted by pointers on each leaf node. IN UBSA, the candidate solution is based on tree topology however, a neighbor generator is used to find better configurations. The acceptance criteria is designed in

such a way that a new configuration is chosen as long as the cost of new configuration is lesser than the current one. UBGGA was found to perform better in terms of resource utilization and energy consumption as it can reduce the number of physical machines used.

A meta-heuristic approach using PSO is introduced in [22]. They have designed and implemented Adaptive Power-Aware Virtual Machine Provisioner (APA-VMP) to handle the resource allocation for workloads in a dynamic environment. The algorithm also considers the power consumption of active servers and the servers which are idle by considering different sleep states to make decisions without degrading performance. They have proved the performance to be better than the previously introduced PSO techniques as the algorithm is self-adaptive and upholds the fitness value corresponding to an optimal solution. The thesis in [23] proposed a novel VM placement algorithm based on ACO meta-heuristic. They show that their algorithm outperforms the First-Fit Decreasing greedy approach as their approach required fewer servers. To achieve both scalability and high data center utilization, a consolidation algorithm based on ACO was also proposed in this thesis. However, the work considers an aggregated resource utilization which might not be suitable to take decisions on dispatching the right VMs. Similar work is done in [24] where ACO based VM placement and VM consolidation algorithms are developed to ensure energy efficiency while supporting QoS and SLA requirements in data centers. The proposed multi-objective ACO algorithm proposed in [23] achieved the best performance in terms of energy efficiency. But in terms of minimizing the number of servers First fit decreasing algorithm showed the best results.

2.3. Deterministic approaches

Deterministic Optimization approaches focus on finding global solution by taking the advantage of the analytical properties of a problem [25]. The quality of the obtained solutions are guaranteed unlike the heuristic methods but it takes immense time to compute the results for a large optimization problem. Constraint Programming approach is being used to address the VM packing problem in [26]. The proposed approach takes into consideration the energy consumption and SLA requirements in combination with utility functions. The authors argue that the Constraint programming approach has many advantages over heuristics-based approaches in terms of generating optimal solutions. The paper proposed in [27] discusses a flexible energy-aware VM allocation and consolidation in a federated cloud data

center. To compute the energy-aware placement of VMs, Constraint Programming (CP) paradigm and the Entropy open source library is used. SLAs are translated in the form of constraints. Optimizer which is the core of the framework works on CP engine and takes SLA constraints and Power Objective function as inputs. The tests conducted showed that a notable amount of energy can be saved with this approach while minimizing CO2 emissions. Linear programming technique and integer optimizers are used in [28] to optimize the management of grid-based datacenter. The problem to be solved is to assign a set of jobs to a set of resources. The scheduling approach makes use of the consolidation methods to find the near-optimal solution which is economically benefiting, results in lower violation of QoS and minimum power usage. The proposed methodology uses migration and turns-off unused physical machines thereby minimizing the energy consumption. Integer linear program is considered to be computationally hard. Near-optimal solutions can be obtained through this approach by understanding the system and having a clear problem statement with good constraints.

2.4. Hybrid approaches

In this paper, the term hybrid approach is referred to the solution approach which makes use of multiple allocation policies for VM consolidation. Hybrid approach can be a combination of heuristic algorithms, meta-heuristic algorithms or both. In [29], the authors have proposed a neural network-based adaptive selector system which can automatically choose appropriate algorithms for dynamic VM consolidation. The hybrid approach incorporates several existing algorithms to choose from. The selection is based on the current environment and cloud provider's priority on energy and SLA violation. Although the proposed idea showed good results in reducing the SLA violations and energy consumption it did not exhibit great improvement in terms of energy consumption. The author describes that this approach would require huge resources and time to train the data-set.

[30] proposes a novel approach for load management using three different existing methods. The best features of ACO, Min-Max Ant System as well as GA algorithm are used to develop a hybrid load management algorithm which reduces SLA violations and number of VM migrations. The proposed work is split into three sections. After the initial allocation, ACO algorithm is applied to obtain pheromone solution. Then, min-max algorithm is run on this data to find the solution which exhibits minimum change. Finally, the result obtained from min-max algorithm is given as input

to the genetic algorithm. From a large pool of solutions, only the optimal solution will be obtained. The fitness function is calculated based on the cost of host and total number of hosts. Overall, this approach provided better results for SLA violations and number of VM migrations when compared to traditional load management techniques. There are many solutions for efficient resource utilization, power management, reduction of SLA violations and dynamic reconfiguration of VMs using various VM placement and migration techniques [31]. However, most of these papers focus on specific problems. More often only the CPU utilization is considered to be the most important resource during dynamic VM consolidation. For an optimal solution, metrics such as energy efficiency, reduction in SLA violations and number of VM migrations should be considered together while taking into account various resources of a physical machine that consumes high power, keeping in mind the dynamic nature of workloads and changing application requirements in a cloud environment [32]. The survey conducted in [33, 34, 29] also supports our

hypothesis that there is a need for hybrid optimization technique which will make use of a combination of various server consolidation methods and finds the best trade off between energy consumption and quality of service. One of the main goals of this research is to address the problem of VM consolidation by taking into consideration multiple objective values since most of the real world problem are characterized by several optimization criteria.

3. Design and Implementation of the conceptual model

This section is dedicated to present the proposed conceptual model of the solution approach for the management of data centre operations. The theoretical framework of the proposed solution approach and the design of evaluation model is presented in Figure 1.

A hybrid approach which uses both heuristic and meta-heuristic approach is proposed to optimize the virtual machine placement considering effective utilization of resources. The conceptual model of this paper is based on our previous work in

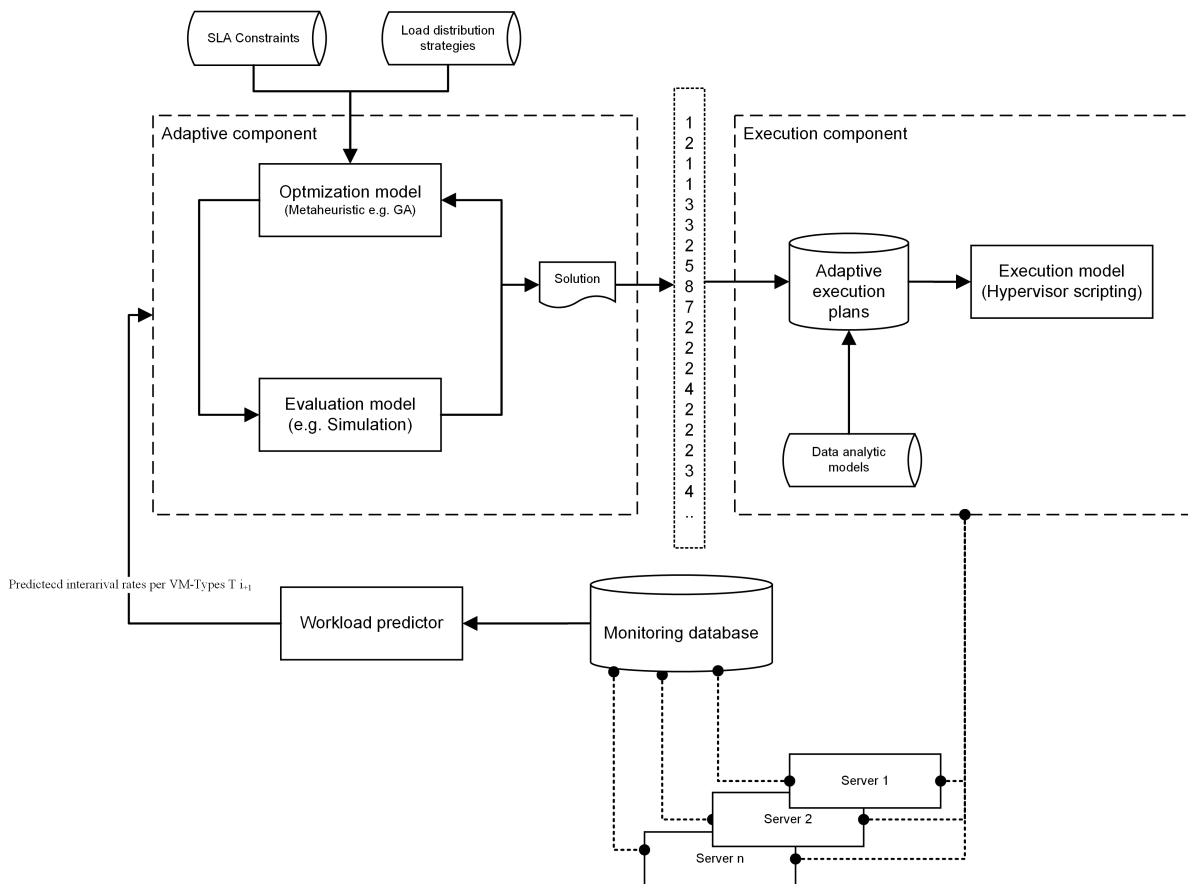


Figure 1. Conceptual model of hybrid load management strategy

[35]. The previous work was an initial analysis in which small scale experiments were conducted in a closed environment to establish the aforementioned hypothesis. Only the number of migrated VMs were considered in the performance metric and reducing energy consumption was not the scope of the work. In this paper, we will be conducting experiments in a large set up considering a heterogeneous environment. Problem instances from a well known benchmark called PlanetLab will be considered and evaluation of our solution approach will be done using heuristics from the state of the art publications.

The proposed framework consists of three main components: the adaptive component (AD), the execution component (AI) and the workload monitoring and prediction component (MP). The MP component will be used to predict the type of workload coming into the system for a specific time span. The data collected regarding the arrival rates of different VM types in this component is passed as one of the inputs to the AD component. The AD component involves the optimization model and the evaluation model. The scope of this paper lies in designing and developing this component to prove the stated hypothesis. The solutions from the AD component that are obtained through a number of simulations will be passed on to the execution component. By passing these solutions to the underlying infrastructure and analyzing the solutions a knowledge base will be created. This component will send continuous feedback to the AD component to control the performance and provide better solutions in each interval. When extensive data is collected and a considerable knowledge is created the AD component can be detached from the system.

3.1. Optimization model

Based on the literature analysis conducted, genetic algorithm is found to be the best approach for VM placement and therefore chosen for the optimization model. The input for the optimization model will include various heuristic based algorithms and SLA constraints. The optimization model will find the best combination of heuristic algorithms based on the inputs provided. The evaluation model will be a simulation model. The heuristic algorithms used for evaluating our solution approach are Power Efficient First-Fit Decreasing (PEFFD), PowerEfficient Best-Fit Decreasing (PEBFD) and Medium-Fit Power Efficient Decreasing (MFPED) from the paper [15]. Modified Best-Fit Decreasing (MBFD) which is proposed for OpenStack Neat [36] is included. Additionally, base algorithms that are available in Cloudsim Plus

simulation framework namely, Best-Fit Static Threshold (BFTHR), Power Aware Best-Fit Decreasing (PABFD), Median Absolute Deviation (MAD) and Static threshold based allocation policy (THR) will be used in our experiments for evaluation. These algorithms are encoded as integers from 1 to 8 in the optimization model. Each individual in the genetic algorithm will contain 24 integers randomly placed which determines the allocation policy which will be changed every hour in a datacenter.

Steps involved in the optimization process

1. The first step in the optimization using genetic algorithm process is to generate an initial population of N solution candidates for evolution. N refers to the population size and $N = 10$ in our set up.
2. A fitness function which is suitable for the problem to be solved should be defined. For every generation, the fitness values of each solution candidate is calculated.
3. Based on their fitness values, elite individuals of a predetermined size are selected and these are directly passed on to the next generation. Elite count is set to 2.
4. Rest of the parent individuals are chosen from the tournament selection. The tournament size is 3.
5. The individuals obtained in step 4 undergo crossover, where the exchange of genes takes place at a random crossover point. If the randomly generated crossover probability is less than the provided crossover rate then the corresponding set of parents undergo crossover. We have set the crossover rate to 0.5.
6. Further, these off-springs will undergo a mutation process where the genes within an offspring are changed based on the type of mutation. If the random mutation probability is less than the given mutation rate then the off-spring will undergo mutation process. The mutation rate is set to 0.2.
7. Every generation's individual is compared with the previous best fitness value for each objective function and if the current individual is better than the previous one, then the best value is updated.
8. The steps from 2 to 7 are repeated until the termination criterion is satisfied. In our algorithm, generation size is set as a termination criteria. which is set to 25.

3.2. The Mathematical Problem Formulation and the Fitness Function

The mathematical model can be defined for the investigated problems as follows.

- Let $P_i = \{P_1, P_2, \dots, P_m\}$ denote a set of physical machines ($i = 1, \dots, m$) in the data center.
- Let $V_i = \{V_1, V_2, \dots, V_n\}$ denote a set of virtual machines ($i = 1, \dots, n$).
- Let $T_i = \{T_1, T_2, \dots, T_o\}$ denote every hour of the day ($0 = 1, \dots, 24$) wherein a different allocation policy is assigned to the datacenter.
- Let $A_i = \{A_1, A_2, \dots, A_r\}$ denote the encoded values of various VM allocation algorithms ($r = 1, \dots, 8$) which are assigned for every T_i to the datacenter based on the optimization model.
- Let $I_i = \{I_1, I_2, \dots, I_s\}$ be a set of individuals in the population ($s = 1, \dots, 10$) of solution candidates.
- Let $C_i = \{C_1, C_2, \dots, C_t\}$ represent a set of chromosomes ($t = 1, \dots, 10$) corresponding to each individual I_i . The length of each chromosome is 24.

As previously mentioned, this work concentrates on targeting multiple optimization criteria. Let S denotes possible solutions in which the allocation policies can be assigned to the system for every hour T_i . Based on the mathematical formulation, it is desired to find a solution candidate $s \in S$ which will contain a random set of encoded values A_i . This solution candidate is subject to minimizing the three objective values as discussed below.

The objective of the proposed solution is to reduce energy consumption, minimize SLA violations and also reduce the number of VM migrations in the datacenter. Therefore, the Fitness function is designed to minimize these three objectives. A solution candidate may not reduce all the three values together, hence there might be several solutions which is best for solving each objective. We consider the fitness values of each solution candidate and the one with the least value is considered to be the best possible solution. The mathematical model for the three objective functions can be described as below.

$$f_1(x) = \sum_{i=1}^N Power_c * Util_p$$

where $Power_c$ is the Power consumed by the host and $Util_p$ is the utilization history of the host in percentage.

SLA violation when the host is not available due to overloading:

$$f_2(x) = \frac{1}{N} \sum_{i=1}^N \frac{R_{o_i}}{R_{a_i}}$$

N is the number of hosts - R_{o_i} is the amount of VM resources that is not allocated leading to an SLA violation. R_{o_i} is counted when the CPU capacity requested exceeds the available capacity. R_{a_i} is the total amount of resources that is requested.

$$f_3(x) = VM_{Migrations}$$

- Corresponds to total number of VM migrations in the datacenter in one day

The multi-objective fitness function is formulated using weighted sum method. Each objective is given a weight based on the priority. If an objective function is given more weight it means that the particular function has higher priority in comparison to other functions [37]. The weight given has a huge influence on the solution that will be obtained at the end of optimization. The obtained objective values are normalized to values between 0 and 1. The dimensionless values are then added up to generate the overall fitness value. The fitness function used in our algorithm can be expressed as the following.

$$F(x) = W_1 f_1(x) + W_2 f_2(x) + W_3 f_3(x)$$

where W_1, W_2 and W_3 are weighting coefficients which satisfies the condition $\sum_{i=0}^n W_i = 1, W_i \in (0, 1)$

4. Experiments and results

In this section, we will describe the experimental set up used for the verification. The computational results obtained for each heuristic solution approaches and the proposed hybrid approach will be demonstrated. The evaluation will be closed by highlighting the findings of the conducted experiments.

4.1. Experimental setup

To evaluate the solution approach, simulation method is adopted. For this, we have used the Cloudsim Plus simulation framework. CloudSim Plus is a popular Java based open source framework which allows modeling and simulation of cloud infrastructure and services. It allows simulating very realistic cloud scenarios using a few lines of code.

Planetlab set up from cloudsim which is originally used in [36] and later in [15] has been adopted. We have considered the infrastructure with heterogeneous set up which is more realistic to a cloud environment. A total of 560 hosts of four different types are included

in this scenario. HP ProLiant ML110 G4 (2 X 1800 MIPS), HP ProLiant ML110 G5 (2 X 2660 MIPS), IBM Xeon X3470 (4 X 2933 MIPS) and IBM Xeon X3480 (4 X 3067 MIPS) are the host types. The experiments are conducted using real cloud workloads obtained for 10 random days from PlanetLab. PlanetLab is a well known benchmark used in a lot of research in the recent years for evaluating their algorithms [8, 11, 15, 29, 36]. Each PlanetLab trace file available, contains CPU utilization measured at every 5 minutes (300 seconds) inside PlanetLab VMs.

We tried the simulation in Cloudsim using the allocation policies from [15] and obtained results which were closer to the ones shown in the paper. However, due to the limitation of the framework, we were unable to conduct our experiments in Cloudsim. These algorithms are used in Cloudsim Plus without any changes to evaluate our proposed approach.

4.2. Results and analysis

The experiments are designed to investigate the heuristics and the hybrid approach based on the objective functions. The weighting coefficients are set to $W_1=0.6$, $W_2=0.2$ and $W_3=0.2$. The used weights are set based on careful analysis of the considered objective values. We tried various combinations of the weights to achieve an acceptable trade-off between minimizing the overall energy consumption and the violations in the SLA as our main requirement. Our analysis suggests that increasing the weights of the SLA and the number of migrated virtual machines minimally deliver better results for both values but strongly impact the minimization of energy consumption. The parameters used to conduct the experiments are shown in Table 1. This set of parameters is based on the analysis presented in [15]. The main idea is to achieve an identical ground for comparison to study the performance of the proposed framework against the other heuristics that are presented in [15] and [36].

Table 1. Parameters used in VM allocation policies

CPU lower threshold	30%
CPU upper threshold	70%
Safety Parameter Linear Regression	1.2

Table 2 shows the average computational result obtained for all 10 problem instances. In terms of energy consumption, our approach has shown considerably major improvement as presented in Figure 2. The proposed genetic algorithm based approach is 30.77% more power efficient in comparison to the PABFD algorithm. In comparison to the other approaches such

as PEBFD, PEFFD and MFPED proposed in [15], the hybrid solution approach is 47.74%, 44.37% and 45.39% better performing. We see that there is a 0.04 units increase in the SLA violation in comparison to PABFD as shown in Figure 3. The number of VM migrations are also reduced from 739 in PABFD to 718 in the hybrid approach as depicted in Figure 4. The proposed approach is about 65.05% energy efficient when compared to the basic algorithms in Cloudsim Plus.

Table 2. Average performance of the proposed hybrid solution

Allocation Policy	Energy Consumption(kWh)	SLAV(%)	#VM Migrations
MAD	912.83	0.02	176
THR	912.75	0.02	214
MBFD	902.62	0.00	241
PEBFD	610.58	0.09	554
PEFFD	573.56	0.10	569
MFPED	584.26	0.10	620
PABFD	460.89	0.12	739
BFTHR	833.97	0.00	0
GA	319.06	0.16	718

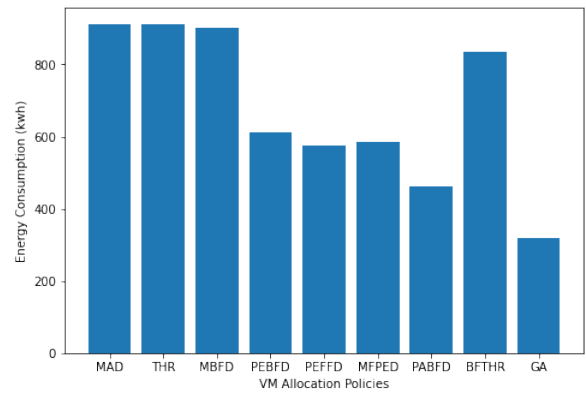


Figure 2. Average energy consumption in the datacenter

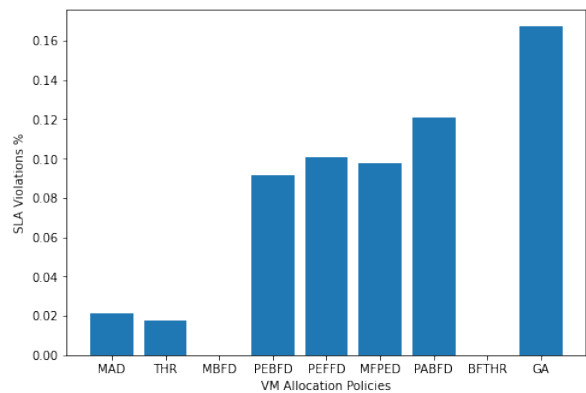


Figure 3. Average SLA violations in the datacenter

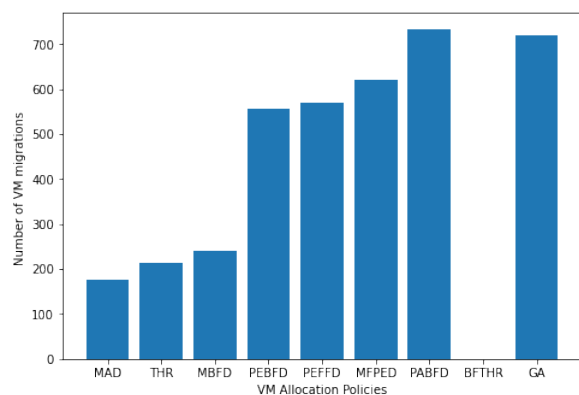


Figure 4. Average number of migrations of virtual machines

The Best-Fit based heuristic algorithms denoted as MBFD and BFTHR show no violations in SLA. However, these algorithms consume high energy as shown in Figure 2. The PABFD algorithm which consumes less power among the heuristic based approaches shows a SLA violation of 0.12% while our approach shows a violation of 0.16%. The SLA violation for all the considered algorithm is however low and the increase of 0.04% in the genetic algorithm can be considered to be a smaller deviation. In terms of VM migrations, the BFTHR did not migrate any VMs. Genetic algorithm initiated 15 migrations less when compared to the PABFD algorithm and 163, 148 and 97 migrations more than that of PEBFD, PEFFD and MFPED respectively as presented in Figure 4.

5. Conclusion and future work

From the experiments conducted using the simulation framework, we can conclude that the hybrid approach outperforms the heuristic based approaches in terms of energy consumption while maintaining a acceptable quality of service and robustness of the system. This means, reduction in the operational costs and minimization of carbon emission is guaranteed by using multiple heuristic algorithms in a adaptive solution approach. We found that the optimization model tries to find a trade-off between energy consumption, SLA violations and number of VM migrations to produce a near to optimal solution. The computational effort is comparatively high but by using high computing distributed systems, the optimization can be run in a couple of minutes. There are several areas which can be looked into in the future. The solution approach can be extended for multiple users by using the concept of adaptive thresholds. The proposed approach enables provisioning of services

based on the cloud user's requirement. In addition, the proposed strategy can be extended to include servers with different power models taking into account the type of the used energy source for operation (e.g. fossil energy, renewable energy). This allows for instance, migrating computationally expensive VMs to servers that are located in data centers with a higher share of used renewable energy for operation, which leads to reducing carbon emission. In the optimization model, different weights can be provided for the objective function based on the individual obligation of each user. Another plan into the future would be to include load management algorithms which take into consideration other resources such as memory and bandwidth in the optimization model and these parameters should be included in the calculation of power consumption. In addition, Implementing the proposed hybrid approach for VM consolidation in a real cloud set up such as Proxmox or Openstack would be an interesting task in the future.

References

- [1] S. S. Gill and R. Buyya, "A taxonomy and future directions for sustainable cloud computing," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–33, 2018.
- [2] S. David, "'hype cycle for cloud computing, 2012'." <https://www.gartner.com/en/documents/2102116>, (accessed on: 14 December 2019).
- [3] M. Attaran, "Cloud computing technology: Leveraging the power of the internet to improve business performance," *Journal of International Technology and Information Management*, vol. 26, pp. 112–137, 08 2017.
- [4] M. C. avdar, "A utilization based genetic algorithm for virtual machine placement in cloud computing systems," 2016.
- [5] A. Andrae and T. Edler, "On global electricity usage of communication technology: Trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, 2015.
- [6] J. Berral, I. Goiri, R. Nou, F. Juli, J. O. Fit, J. Guitart, R. Gavald, and J. Torres, *Toward Energy-Aware Scheduling Using Machine Learning*, pp. 215–244. 07 2012.
- [7] A. Nahhas, S. Bosse, D. Staegemann, M. Volk, and K. Turowski, "A holistic view of the server consolidation and virtual machines placement problems," 11 2019.
- [8] S. Hasan and E.-n. Huh, "Heuristic based energy-aware resource allocation by dynamic consolidation of virtual machines in cloud data center," *KSII Transactions on Internet and Information Systems*, vol. 7, pp. 1825–1842, 2013.
- [9] S. S. Skiena, *The Algorithm Design Manual*. London: Springer, 2008.
- [10] J. Xu and J. A. B. Fortes, "Multi-objective virtual machine placement in virtualized data center environments," in *2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing*, pp. 179–188, IEEE, 18/12/2010 - 20/12/2010.

- [11] Anton Beloglazov, Jemal H. Abawajy, and Rajkumar Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing," *Future Generation Comp. Syst.*, vol. 28, pp. 755–768, 2012.
- [12] C. Chang, S.-L. Peng, Y.-H. Liao, and R.-S. Chang, "Green computing: An sla-based energy-aware methodology for data centers," *Frontiers in Artificial Intelligence and Applications*, vol. 274, pp. 1345–1354, 01 2015.
- [13] V. Seema, P. Ritesh, P. Hetal, and P. Sandipkumar, *Efficient Virtual Machine Management Based on Dynamic Workload in Cloud Computing Environment*. Piscataway, NJ: IEEE International Advance Computing Conference, 2015.
- [14] A. Verma, P. Ahuja, and A. Neogi, "pmapper: Power and migration cost aware application placement in virtualized systems," vol. 5346, pp. 243–264, 2008.
- [15] F. F. Moges and S. L. Abebe, "Energy-aware vm placement algorithms for the openstack neat consolidation framework," *Journal of Cloud Computing*, vol. 8, no. 1, p. 205, 2019.
- [16] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. Cambridge, MA, USA: MIT Press, 1992.
- [17] S. Kirkpatrick, C. Gelatt, and M. Vecchi, "Optimization by simulated annealing," *Science (New York, N.Y.)*, vol. 220, pp. 671–80, 06 1983.
- [18] M. Dorigo, M. Birattari, and T. Sttze, "Ant colony optimization," *Computational Intelligence Magazine, IEEE*, vol. 1, pp. 28–39, 12 2006.
- [19] J. Kennedy, *Particle Swarm Optimization*, pp. 760–766. Boston, MA: Springer US, 2010.
- [20] D. U. Chandrasekaran K, "Load balancing of virtual machine resources in cloud using genetic algorithm," 2013.
- [21] V. P. Anuradha and D. Sumathi, "A survey on resource allocation strategies in cloud computing," *2014 International Conference on Information Communication and Embedded Systems, ICICES 2014*, 2015.
- [22] R. Jeyarani, N. Nagaveni, and R. Vasanth Ram, "Design and implementation of adaptive power-aware virtual machine provisioner (apa-vm) using swarm intelligence," *Future Generation Computer Systems*, vol. 28, no. 5, pp. 811–821, 2012.
- [23] E. Feller, *Autonomic and Energy-Efficient Management of Large-Scale Virtualized Data Centers*. Distributed, parallel, and cluster computing, Université Rennes, 2012.
- [24] 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," *Sustainable Computing: Informatics and Systems*, vol. 17, pp. 9–24, 2018.
- [25] M.-H. Lin, J.-F. Tsai, and C.-S. Yu, "A review of deterministic optimization methods in engineering and management," *Mathematical Problems in Engineering*, vol. 2012, 06 2012.
- [26] Hien Nguyen Van, Frédéric Dang Tran, and Jean-Marc Menaud, "Autonomic virtual resource management for service hosting platforms," *Workshop on Software Engineering Challenges in Cloud*, pp. 1–8, 2009.
- [27] C. Dupont, T. Schulze, G. Giuliani, A. Somov, and F. Hermenier, "An energy aware framework for virtual machine placement in cloud federated data centres," *2012 Third International Conference on Future Systems: Where Energy, Computing and Communication Meet (e-Energy)*, pp. 1–10, 2012.
- [28] J. T. Josep Ll. Berral, Ricard Gavaldà, "An integer linear programming representation for datacenter power-aware management," 2010.
- [29] J. Witanto, H. Lim, and M. Atiqzaman, "Adaptive selection of dynamic vm consolidation algorithm using neural network for cloud resource management," *Future Generation Computer Systems*, vol. 87, 05 2018.
- [30] K. Kaur and A. Kaur, "A hybrid approach of load balancing through vms using aco, minmax and genetic algorithm," in *2016 2nd International Conference on Next Generation Computing Technologies (NGCT)*, pp. 615–620, Oct 2016.
- [31] Damien Borgetto, Michael Maurer, Georges Da-Costa, Jean-Marc Pierson, and Ivona Brandic, "Where energy, computing and communication meet (e-energy), 2012: Third international conference on future energy systems ; 9 - 11 may 2012, leganés (madrid), spain," 2012.
- [32] Z. Zhou, J. Abawajy, M. Chowdhury, Z. Hu, K. Li, H. Cheng, A. A. Alelaiwi, and F. Li, "Minimizing sla violation and power consumption in cloud data centers using adaptive energy-aware algorithms," *Future Generation Computer Systems*, vol. 86, pp. 836–850, 2018.
- [33] Z. Usmani and S. Singh, "A survey of virtual machine placement techniques in a cloud data center," *Procedia Computer Science*, vol. 78, pp. 491–498, 2016.
- [34] S. Challita, F. Paraiso, and P. Merle, "A study of virtual machine placement optimization in data centers," 04 2017.
- [35] A. Nahhas, S. Bosse, M. Pohl, and K. Turowski, "Toward an autonomic and adaptive load management strategy for reducing energy consumption under performance constraints in data centers," in *Proceedings of the 9th International Conference on Cloud Computing and Services Science*, pp. 471–478, SCITEPRESS - Science and Technology Publications, 02/05/2019 - 04/05/2019.
- [36] A. Beloglazov and R. Buyya, "Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers," *Concurrency and Computation: Practice and Experience*, vol. 24, no. 13, pp. 1397–1420, 2012.
- [37] N. Gunantara, "A review of multi-objective optimization: Methods and its applications," *Cogent Engineering*, vol. 5, no. 1, 2018.