Hindawi
Wireless Communications and Mobile Computing
Volume 2023, Article ID 9856960, 1 page
https://doi.org/10.1155/2023/9856960

Retraction

Retracted: Cloud-IoT Resource Management Based on Artificial Intelligence for Energy Reduction

Wireless Communications and Mobile Computing

Received 8 August 2023; Accepted 8 August 2023; Published 9 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] W. B. Daoud, A. Mchergui, T. Moulahi, and A. Alabdulatif, "Cloud-IoT Resource Management Based on Artificial Intelligence for Energy Reduction," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 2248962, 8 pages, 2022.

Hindawi
Wireless Communications and Mobile Computing
Volume 2022, Article ID 2248962, 8 pages
https://doi.org/10.1155/2022/2248962

Research Article

Cloud-IoT Resource Management Based on Artificial Intelligence for Energy Reduction

Wided Ben Daoud,[1] Abir Mchergui ,[2] Tarek Moulahi,[3] and Abdulatif Alabdulatif [4]

[1]*NTS'Com Research Unit, ENET'Com, University of Sfax, Sfax, Tunisia*
[2]*Department of Computer Science, ISG Gabes University, Tunisia*
[3]*Department of Information Technology, College of Computer, Qassim University, Buraydah, Saudi Arabia*
[4]*Department of Computer Science, College of Computer, Qassim University, Buraydah, Saudi Arabia*

Correspondence should be addressed to Abir Mchergui; abir.mchirgui1437@gmail.com
and Abdulatif Alabdulatif; ab.alabdulatif@qu.edu.sa

Received 4 February 2022; Revised 28 February 2022; Accepted 5 March 2022; Published 25 March 2022

Academic Editor: Rashid A Saeed

The rapid growth in demand for cloud services has led to the creation of large-scale data centers, which allows application service providers to lease data center capacity for application deployment as per user requirement in terms of quality of services (QoS). These data centers consume a lot of electrical power which contributes to increased operating costs and carbon dioxide emissions. In addition, modern cloud computing environments must provide QoS for their customers which leads them to a need to make a power-performance compromise that is to say in terms of energy consumption and service-level agreement (SLA) compliance. That is why, we introduce, in this paper, an intelligent resource management policy for cloud data centers. The goal is to dynamically allocate and continuously consolidate virtual machines taking advantage of live migration and disengage inactive nodes to minimize power feeding in this cloud environment while maintaining the quality of service. We integrate some artificial intelligence concepts to ensure a dynamic resource management and a better power-performance compromise and significantly reduce the consumed energy.

1. Introduction

In recent years, the world of telecommunication has been perceiving the emergence of cloud computing, allowing users to outsource their applications and exploit IT resources through the Internet, without having to manage subinfrastructure underlying, often complex. Therefore, the cloud, as shown in Figure 1, consumes a large amount of energy to provide efficient and reliable services to users [1]. In 2014, data centers used nearly 1.62% of the energy consumed in the world, while the consumption for the year 2020 is of the order of 140 billion KWh. In addition, the consumption of these data centers around the world is expected to double every five years, which can generate a huge cost to business and enterprises. Indeed, these data centers are responsible for emitting 2% of CO_2 into the atmosphere [2]. The ever-increasing demand for cloud services and the desire to provide a certain quality of services require providers to invest large amounts of capital in order to multiply their hosting offers in several geographical areas. With this large-scale deployment of huge data centers, the energy consumption of the cloud increases accordingly [3, 4]. From a 2011 Digital Power Group report, scientist said that if the cloud were a country, it would be the world's 5th largest electricity consumer, about 750 trillion kWh, all in just one year.

A high cost associated to electricity consumption causes service providers to work better in order to solve this problem. Therefore, it is important to work on the source of energy consumption of the cloud, taking into account the technologies that guarantee its proper functioning in order to meet the users' needs, at all times, without interruption or degradation of the quality of services (QoS) [5, 6].

Cloud data centers have become overwhelmed with data-intensive applications due to the limited computational capabilities of mobile terminals. So, cloud computing becomes a backbone in the various fields of industry and Academia for providing the storage backups and load balancing and most importantly providing the dynamic resource scheduling on a real-time basis. Therefore, energy consumption will be very high by the various cloud data centers which lead to a very high amount of operational cost and harm the environment as well. Compute nodes are usually distributed in edge environments, enabling crucially efficient task scheduling among those nodes to achieve reduced processing time. Moreover, it is imperative to conserve edge server energy, enhancing their lifetimes. Hence, various energy-efficient algorithms have been proposed by the researchers which reduce the energy efficiency of cloud environment. But all these algorithms have been evaluated using the same experimental environment which gives the related results, and it becomes difficult for the researchers to choose the best algorithm among them. In this work, we are interested in minimizing energy consumption in the cloud, using an efficient and adequate method. Our approach is based on reinforcement learning (Q-learning), which is a machine learning technique that consists of letting the algorithm learn from repeated scenarios following the decision-making it has to make. Moreover, the proposed solution based on the use of dynamic allocation policy of virtual machines allows to make a better choice of VMs' location on physical machines (hosts) [7]. Indeed, we offer an autonomous VM consolidation method that includes an intelligent RL learning agent to optimize the allocation of VMs around the data center in the aim of enhancing energy economy while simultaneously offering increased performance.

2. Related Works

With the fast expansion of distributed cloud computing network services, the scale of data in various disciplines, such as scientific computing, data processing, bioinformatics, and Internet of Things (IoT) operations, has expanded. Millions of operations are done in cloud data centers by thousands of high-performance servers deployed [8, 9]. The cloud provides a range of services via virtual machines (VMs), and virtualization is one of the most advantageous features for consumers who may take use of these many kinds of services. These VMs typically require a lot of energy. Such energy use increased power costs and has a negative environmental impact in [3]. Indeed, energy efficiency is a big concern for cloud services providers and users. For that, researchers have attempted to suggest solutions to solve this issue. In this section, some of these important research articles, in the literature, relating this problem are covered. It is projected that the data center emits 62 million tons of carbon dioxide (CO_2) into the environment [10], despite several efforts to reduce energy usage by various technologies. This research looked at a number of independent tasks using quality of service criteria. Consequently, the selection of the appropriate virtual machine from a heterogeneous resource setup with the least amount of energy consumption and a task's satisfied QoS is becoming a challenging issue. Therefore, there is a massive waste of resources and an increase in energy usage. Despite this, it is extremely beneficial to schedule all jobs to appropriate servers in order to achieve optimal energy usage. Many researchers devised these challenges to that end. For example, one research [11] looked at the energy-efficient scheduling challenge in heterogeneous devices. The objective was to schedule all jobs in accordance with their QoS needs while consuming the least amount of electricity. [11] demonstrates that hybrid data centers constructed with higher efficiency servers may save electricity. The issue of energy usage is virtually inextricably connected to the processing duration of the activity. This presents a number of problems in determining the proper operating time for activities. Energy savings in virtualized data centers have piqued the interest of academics and business alike. Effective scheduling strategies and multicomponent systems have been widely researched in [12]. Qin et al. [12] use the intratask DVFS scheduling technique, based on task profile information, under the assumption of zero transmission latency, expressed as an ILP model, to better optimize the energy for real-time activities as systems need increasing power. The major objective of this ILP model is to find the optimal processing frequency for each basic block in order to obtain the lowest average energy. To compensate the DVFS conversion overhead, the ILP method is modified to calculate the optimal execution frequency while also determining the appropriate program places to insert the conversion instructions. Wu et al. presented an energy-aware task scheduling algorithm (ETSA) in [3] to overcome the drawbacks associated with task consolidation and scheduling. To make a scheduling choice, the suggested ETSA technique improves task completion time and overall resource usage and proposes a normalization method.

Traditional techniques may not necessarily result in a suitable timetable. The majority of research concentrate on heuristic algorithms, which are often based on greedy local optimum selection heuristics. Nonetheless, due to their flexibility, numerous well-known metaheuristic strategies, such as ant colony optimization (ACO), particle swarm optimization (PSO), and genetic algorithm (GA), have proven prominent in task scheduling issues. The authors of [2] suggested a two-stage energy and performance-efficient task scheduling algorithm (EPETS). The first step of scheduling aided in minimizing processing time and meeting job deadlines while ignoring energy usage. The second step of task reassignment scheduling was to locate the optimum execution site while staying within the timeline restriction and consuming the least amount of energy. Recently, the study of energy conservation strategies for virtual machine (VM) configuration, migration, and consolidation has become a research issue. In fact, several techniques for solving the VM consolidation issue utilizing evolutionary algorithms have been presented. In this context, [13] proposed a VM placement method that uses a priority-based probability scheduling model to evaluate resources, virtual machine state, QoS metrics, and I/O data. Data location is evaluated during the VM placement step to minimize needless

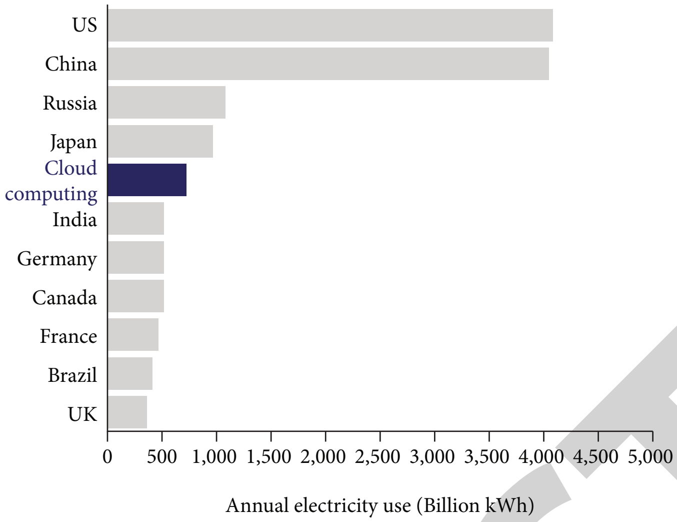

Figure 1: Annual electricity use.

migration. The ant colony optimization (ACO) strategy has been utilized to optimize various objectives in numerous VM consolidation approaches [14]. Terra-Neves et al. [15] created an improved discrete differential evolution algorithm to solve the VM placement problem. The authors attempted to reduce energy usage and overloading danger. The authors of [16] tackled the issue of reducing needless VM migrations by proposing a discrete-time Markov chain (DTMC) model to forecast future resource use. Then, utilizing the e-dominance-based multiobjective artificial bee colony (e-MOABC) method, a multiobjective VM placement strategy is presented to accomplish the optimal VMs to PM mapping. To fulfill SLA and QoS criteria, the proposal may efficiently manage total energy usage, resource waste, and system dependability. The primary limitation of most current research is that they mainly concentrated on lowering the number of active PMs utilizing VM live migration to avoid inefficient resource consumption. Although these approaches are extremely efficient in terms of energy management, they overlook the detrimental impact of frequent VM consolidation on system reliability. As a result, throughout the decision-making process, a holistic perspective and the examination of numerous aspects are critical. The sleep mode-based energy saving method is implemented by putting the idle server to sleep in a low-power state. Jin et al. presented a clustered VM allocation approach on the cloud system's resource layer that is based on a sleep mode with a wake-up threshold. They calculated the system's performance metrics in terms of average request delay and energy saving rate by constructing a queue with an N-policy and asynchronous vacations of partial servers [17]. A multitier cloud architecture is made up of many distinct layers, including "application layer," "management layer," and "resource layer" [18]. The DVFS technique is a technique that acts on the frequency and voltage's CPU of each server in the cloud in order to reduce their energy consumptions. The execution of a request requires the use of the CPU, memory, and disk. It is advantageous to reduce the frequency of the CPU to execute a request that uses a small proportion of the CPU and a high proportion of the memory at the server level [19]. Indeed, if the CPU usage rate for a request is high, then the execution time of this request will be low. However, if we reduce the CPU frequency, the execution time will increase, because they are inversely proportional. As a result, the cloud performance will be degraded. Therefore, the power consumption can be reduced by lowering the CPU frequency during the processor wait times. In this context, Ghribi proposed a DVFS technique based on query execution statistics and a learning algorithm to select the optimal value of the voltage/frequency ratio for query execution. Indeed, this algorithm can predict the optimal value of the voltage/frequency ratio from query execution statistics, such as the number of executed statements and the number of failed statements and clock cycles [20]. This technique can greatly reduce the power consumption of the CPU. However, the application of a change of frequency is not instantaneous, and the change from one frequency to another takes a little time, which may slow down the execution of the application, which will undoubtedly cause an overconsumption of energy. In Table 1, we elaborate and summarize a comparison between these related works based on the used algorithm, the drawbacks, and the tolls used for experimentation.

3. Overview of our Contribution

In this work, we propose dynamic allocation through the migration of virtual machines according to current resource needs and their availability. The purpose of the reallocation is to minimize the number of physical nodes serving as the current workload, while inactive nodes are turned off to reduce power consumption. The allocation of virtual machines is to place virtual machines on hosts in a methodical way and to be efficient so that they will meet the requirements of the cloud because they host a large number of services. With this in mind, the dynamic allocation of VMs will help to reduce energy consumption in this environment.

Table 1: Method comparison.

Ref.	Used algorithm	Drawbacks	Tools used for experimentation
[6]	VM scheduling technique	Problems in determining the proper operating time for activities.	CloudSim
[7]	Intratask DVFS scheduling technique		CloudSim
[3]	Energy-aware task scheduling algorithm(ETSA)	Not necessarily result in a suitable timetable	CloudSim
[11]	Two-stage energy and performance-efficient task scheduling algorithm (EPETS)	Problems in SLA violation and provided scalability	Scheduler is implemented
[12]	Priority-based probability scheduling model	Address more objectives such as resource overcommitment and bandwidth resource constraints	Simulator (self-designed)
[22]	ACO	System reliability problem	Simulator (self-designed)
[2]	An improved discrete differential evolution algorithm	Slow down the execution of the application	CloudSim
[13]	Discrete-time Markov chain (DTMC) model	Overlook the detrimental impact of frequent VM consolidation on system reliability	CloudSim
[14]	Clustered VM allocation approach		CloudSim
[16]	DVFS technique	Slow down the execution of the application	CloudSim
[17]	DVFS technique based on query execution statistics and a learning algorithm	The application of a change of frequency is not instantaneous Slow down the execution of the application	CloudSim

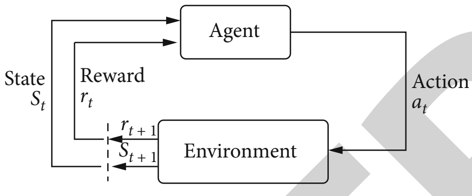

Figure 2: Illustration of the agent-environment interaction system.

3.1. Reinforcing Learning (Q-Learning) Method. In a highly stochastic, nondeterministic environment, reinforcement learning allows an agent to learn optimum activity. A cyclical learning of state-action-reward interactions occurs when an agent interfaces with its environment in order to obtain information about how to improve its behavior and identify the best policy to fulfill its objectives, as shown in Figure 2. Based on trial and error, the agent must determine which acts offer the highest reward. Furthermore, the chosen action has an influence on both the immediate reward and following states and hence all subsequent rewards. The Markov decision process (MDP) is widely regarded as the gold standard for describing learning in sequential decision-making situations with unfavorable uncertainty. Using simulated trials, the MDP model enables agents to acquire an optimal policy progressively [21]. This characteristic asserts that in order to forecast all future states, just the actual state of the environment is needed [7]. The MDP framework includes states, actions, transition probabilities, and rewards, respectively $(S, A, p, \text{and } q)$, wherein

(i) S represents a set of possible states

(ii) A signifies a set of actions

(iii) $p(s_{t+1} \mid s_t, a_t)$ signifies the probability distribution governing state transitions

(iv) $q(s_{t+1} \mid s_t, a_t)$ signifies the probability distribution regulating the rewards received $R(st, at)$

Often, the learning method is divided into distinct time stages t. The learning agent is in state $s \in S$ at the conclusion of each time step t. The agent chooses an action at $\in A(s_t)$, where $A(s_t)$ refers to the collection of potential actions in the current state s. The agent receives a reward $R(st, a_{t+1})$ after completing the specified action, which causes an environmental state change st. Given that the agent is in state s and chooses action at, the state transition probability $p(s_{t+1} \mid s_t, a_t)$ estimates the likelihood of a transition to $st + 1$. The expected reward obtained by the agent after migrating from state st to state s by executing $q(s_{t+1} \mid s_t, a)$ indicates the expected reward received by the agent after transitioning from state st to state s by executing action $A(s_t)$. The update rule is defined as [44]

$$Q(\mathrm{st}, a_t) \longleftarrow Q(\mathrm{st}, a_t) + \alpha[\mathrm{rt} + 1 + \gamma \max Q(\mathrm{st} + 1, a) - Q(\mathrm{st}, a_t)]. \quad (1)$$

The proposed solution is based on reinforcement learning (Q-learning). This method makes it possible to learn a policy based on which action carried out in each state of the environment.

This works by learning a Q-rated action-status function that determines the potential gain, i.e., long-term reward $Q(s, a)$, gotten by the fact of carrying out a certain action in a certain state s by following an optimization policy. When this Q function is known, learned by the agent, the optimal policy can be built by selecting the action at maximum value for each state, that is, by selecting the action a which

```
for each VM in VMigrationList do
for each server in ServerList do calculate hostUsageRate
    end for
     calculate VMSize
    for each host do Actions ← V MSize + hostUsageRate
    end for
    selectserverfromActionsusing$\pi$
    allocateV M
    observe globalState + 1,reward
    calculate Q(s,a) ← Q(s,a) + $\alpha$[r + $\gamma$maxaQ(st + 1,a)Q(s,a)]
      update QValueMatrix
      globalState ← globalState + 1
  end for
```

Algorithm 1: Q-learning VM consolidation algorithm.

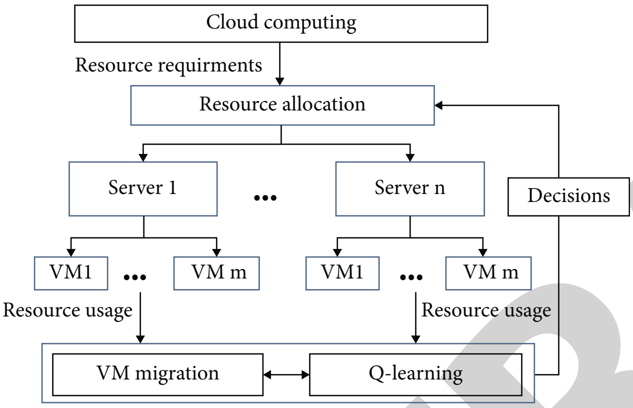

Figure 3: Overview of the proposed method.

maximizes the value $Q(s, a)$ when the agent is in the state s. Reinforcement learning is used to teach an agent how to behave in an environment that may be real or virtual. In fact, the agent learns and makes decisions about the state of an environment that provides the training data. Indeed, the agent receives the reward he deserves after taking the action (a_t) to the state (s_t). The agent's goal is to maximize rewards.

The agent takes actions (activate or put on standby) in the environment at the state s. It tries to decide what action has to be performed on the virtual machines at the state s. There are key moments when making this decision. This moment coincides with the arrival of a request to be processed. The state of the server at time t is St, which can be in standby mode or active mode. The requests that the server must process are fixed, and their processing times differ from one request to another. In this work, we are interested about requests of the same type because not all requests have the same processing time since they are of different types (execution of an application, storage, etc.), and therefore, they do not have the same latency.

3.2. Virtual Machine Migration. Virtual machine migration involves transferring virtual machines from one server to another server. This transfer includes the transfer of memory, CPU, and disk to the destination server. However, the migration process can also be used in an intercloud. As part of our project, we are studying the migration carried out in the same cloud, which makes it possible to solve, among other things, the problem of fault tolerance, or the continuity of services. Migration techniques are classified as cold migration (nonlive migration) and hot migration (live migration). In our work, we focus on live migration because this migration is much more important because it is carried out without interrupting data processing during execution, since the objective is to satisfy users in order to offer them a better quality of services (QoS) in terms of availability of all services offered by cloud computing at any time. This type of migration is done by transferring a running virtual machine from one server to another, maintaining the state of the machine during the transfer process. Thus, the user of the services hosted on the virtual machine is not aware of the change of server that allows to maintain the services of the users during the duration of the process.

3.3. Q-Learning-Combined VM Migration Implementation. In the proposed model, VMs, in the cloud data center, are assigned to a host server based on the required services, which vary considerably and dynamically over time. For the purpose to minimize energy usage and boost QoS, it is necessary to optimize VM distribution across the data center via live migration and reallocate VMs to other servers in the data center. In our work, the system supervises the resource consumption of each server in the data center, continuously. When detecting overloaded hosts, the system selects VMs to be migrated, and these VMs are consolidated onto a more appropriate server while keeping performance in mind. In our system model, we choose the live migration for the VM. It is made up of the VM selection method combined with the intelligent Q-learning algorithm. It makes the final decision on VM migration and allocation. In this way, cloud system will improve energy efficiency. The suggested Q-learning-based RL method, combined with VM migration, is detailed in Algorithm 1. Firstly, the global state of the environment is computed, and a list of all feasible hosts in the data center with accessible resources for running VMs is generated. From the migration list, a VM is chosen. The VM selection technique shown in Figure 3 generates this list. The size of the VM indicated (VMSize) is then computed.

Following this measure and this combination, the VM will be reassigned to the most available and suitable server, and the global state will be recomputed. The system also receives a reward. The *Q*-value update rule is then evaluated, and the result is saved in the *Q*-value matrix. Finally, the global state is refreshed in preparation for the next iteration. This method is repeated until all of the VMs in the migration list have been reassigned to different hosts in the environment. With our best energy management algorithm, we sort all virtual machines in descending order of current use and allocate to a host that provides a small increase in power consumption due to this allowance. This makes it possible to take advantage of the heterogeneity of the physical nodes (servers) by choosing the most energy efficient. The BEM algorithm verifies the availability of resources at the physical machine and distinguishes virtual machines according to their reduction in the capacity of the processor. The virtual machine that has the lowest use of the processor is allocated first and process continues. This proposed algorithm uses the location service of cloud mobile servers. The virtual machine that is close to the server is first evaluated with the distance measurement.

3.4. VM Migration Based on RL. VM consolidation increases resource management and data center optimization by putting a higher number of virtual machines onto a smaller server utilizing live migration to minimize resource use while still meeting user-specified service-level agreements (SLA). Over the years, several learning approaches have been discovered and suggested, in the machine learning (ML) field. Yet, all learning algorithms are created for a specific matter, and we have no global solution for all sorts of problem domains. Hence, a specific system process must be taken based on its appropriateness to the problem [15, 22]. In this paper, we introduce a self-optimizing RL-based VM migration method for optimizing VM allocation and achieving larger energy gains which, thus, guarantees increasing data center service. We use the Q-learning method to solve the VM consolidation issue and assess its effectiveness using a variety of cloud performance indicators. Through repeated interactions with the environment, the RL agent learns an optimal resource allocation strategy based on the present state of the system.

4. Performance Evaluation

This section is dedicated to the implementation and the consolidation phase of virtual machines to reduce energy consumption in cloud. To do this, we carry out simulations in in Java environment, using CloudSim simulator for the purpose of carrying out series of experiments.

4.1. Simulation Parameters. CloudSim is a generalized and extensible simulation framework that enables the modeling, simulation, and experimentation of new cloud infrastructures and associated application services. We used the version of the CloudSim 3.0.3 simulator. The CloudSim simulator consists of several classes forming its constituent blocks. In this part, we conduct two experiments on data centers, which make up the cloud computing environment, with the aim of minimizing its energy consumption without hindering its proper functioning. For that, we are interested in the operation of the IqrMc class. Then, we will implement the virtual machine allocation policy (Iqr) with the BEM algorithm (best energy management). The second step is to implement the virtual machine migration policy in order to reduce the energy consumption of the cloud. This class always applies random policy as a workload. For the simulation, we are interested in the following parameters.

(i) The total time of the simulation is the execution time of all the instructions done by the virtual machines

(ii) The energy consumed is the energy consumed by the physical machines (hosts) during this simulation

(iii) The number of virtual machines migrated is the number of VMs migrated from one physical machine to another during the simulation

(iv) Performance degradation due to migration represents the percentage of performance degradation of physical machines related to migration of VMs

(v) Global violation of SLA is the level of noncompliance with the SLA (service-level agreement) contract recorded at the end of the simulation

(vi) Average SLA violation is the level of noncompliance with the SLA contract, recorded during the simulation

(vii) The number of physical machines turned off represents the total number of physical machines shutdown due to the total migration of VMs on these physical machines (hosts)

4.2. Simulation Results and Discussion Study. After a series of experiments, we can observe the results in Table 2. The primary metric utilized to assess our technique is energy usage. As seen in Table 2, the consolidation through L1 consumes less energy than consolidation via LE. So, we can deduce that our implementation consumes energy, 11.17% less than that consumed in the existing work. On the other hand, our proposal has fewer shutdown servers than LRRMMT. This is due to the fact that there are more moderately loaded servers in LBBMC, which is a positive thing because these active servers require less energy. This will also help the data center to remain balanced for a longer period of time.

Furthermore, the execution time of our technique is longer than that of the LRRMMT method. This is due to the calculation time of the balancing factor F, which is not taken into consideration in LRRMMT, as well as the parameters of our matching technique, which is used for live migration and considers the energy consumption of the VM on the source server as well as the migration energy cost. Table 2 shows that the number of migrations in LBBMC is smaller than in LRRMMT. Because the amount of migrations

TABLE 2: Simulation parameters and results.

Parameters	Proposed contribution	Existent alg
Total time of the simulation	86400 seconds	98250 seconds
Energy consumption	41,63 KWh	76,86 KWh
The number of virtual machines migrated	3841	5085
The decline in SLA performance due to migration	0,24%	0,26%
Overall SLA violation	2,16%	1,13%
Average ALC violation	12,48%	10,81%
The number of physical machines switched off	688	1517

contributes to energy consumption, our reassignment technique prevents unnecessary migrations, which increases the migration overload of VM, while it is better to minimize this overload for the proper functioning of physical machines.

A significant number of the migrated virtual machines contribute to the degradation of SLA performance. These migrated VMs are the base of 0.26% of this degradation, for the existing work, against 0.24% in our implementation. Moreover, the overall SLA violation is 1.13% for existing work and 2.16% after implementing our algorithm, which means that our approach does not comply with all clauses established by cloud providers.

There are 1517 hosts shutdown in the existing work due to a significant migration of VMs, resulting in a degradation of SLA performance, while we recorded 688 hosts that are shut down for our proposal. This implies that the existing work degrades much the performance due to the migration process.

For the same characteristics of a cloud data center, we find that our implementation reduces the energy consumption of the cloud, compared to the existing work, but less respects the established clauses of the SLA contract. Nevertheless, in terms of compromise between the energy consumed and the degradation of the SLA contract, our approach offers a good alternative to remedy this problem, thanks to the efficiency of its good strategy to migrate as few VM as possible, in order to keep the cloud running smoothly.

The obtained results shows that the technique of dynamic allocation and consolidation of virtual machines deactivates inactive physical machines (hosts) and provides energy savings in the cloud computing environment and can be applied in the real world, on real cloud data centers, in order to obtain an optimization of the energy management in these data centers, where the latter will contribute enormously to the reduction of greenhouse gases because it minimizes energy consumption.

However, our proposed system does not consider some other system resources in the reallocation of virtual machines, such as network interface and disk storage, as these resources contribute significantly to the overall energy consumption.

5. Conclusion and Future Work

In this paper, live migration method is combined with the BEM algorithm with the goal of minimizing energy consumption and providing better performance in a cloud environment. The simulation results show that our proposal improves workload management and lowers energy consumption for a large number of servers in a data center.

As a future work, we intend to expand our efforts in order to adapt our algorithm to multiple data centers. Our next goal is to look into new parameters for better energy management in cloud computing.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The researchers would like to thank the Deanship of Scientific Research, Qassim University, for funding the publication of this project.

References

[1] W. Li, H. Cui, T. Nemeth et al., “Cloud-based health-conscious energy management of hybrid battery systems in electric vehicles with deep reinforcement learning,” *Applied Energy*, vol. 293, p. 116977, 2021.

[2] M. Hussain, L. F. Wei, A. Lakhan, S. Wali, S. Ali, and A. Hussain, “Energy and performanceefficient task scheduling in heterogeneous virtualized cloud computing,” *Sustainable Computing: Informatics and Systems*, vol. 30, 2021.

[3] Y. Wu, M. Tornatore, S. Ferdousi, and B. Mukherjee, “Green data center placement in optical cloud networks,” *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 3, pp. 347–357, 2017.

[4] S. Mahfoudhi, M. Frehat, and T. Moulahi, “Enhancing cloud of things performance by avoiding unnecessary data through artificial intelligence tools,” in *2019 15th international wireless communications Mobile computing conference (IWCMC)*, pp. 1463–1467, IEEE, 2019.

[5] X. You, D. Sun, X. Lv, S. Gao, and R. Buyya, “MQDS: an energy saving scheduling strategy with diverse QoS constraints towards reconfigurable cloud storage systems,” *Future Generation Computer Systems*, vol. 129, pp. 252–268, 2022.

[6] F. E. F. Samann, S. R. Zeebaree, and S. Askar, “IoT provisioning QoS based on cloud and fog computing,” *Journal of*

Applied Science and Technology Trends, vol. 2, no. 1, pp. 29–40, 2021.

[7] C. H. Tran, T. K. Bui, and T. V. Pham, “Virtual machine migration policy for multi-tier application in cloud computing based on Q-learning algorithm,” *Computing*, vol. 9, pp. 1–22, 2022.

[8] T. Moulahi, “Joining formal concept analysis to feature extraction for data pruning in cloud of things,” *The Computer Journal*, 2021.

[9] T. Moulahi, S. El Khediri, R. U. Khan, and S. Zidi, “A fog computing data reduce level to enhance the cloud of things performance,” *International Journal of Communication Systems*, vol. 34, no. 9, p. e4812, 2021.

[10] P. Sharma, P. Pegus II, D. Irwin, P. Shenoy, J. Goodhue, and J. Culbert, “Design and operational analysis of a green data center,” *IEEE Internet Computing*, vol. 21, no. 4, pp. 16–24, 2017.

[11] X. Li, W. Yu, R. Ruiz, and J. Zhu, “Energy-aware cloud workflow applications scheduling with geo-distributed data,” *IEEE Transactions on Services Computing*, 2020.

[12] Y. Qin, G. Zeng, R. Kurachi, Y. Li, Y. Matsubara, and H. Takada, “Energy-efficient intratask dvfs scheduling using linear programming formulation,” *IEEE Access*, vol. 7, pp. 30536–30547, 2019.

[13] M. H. Sayadnavard, A. T. Haghighat, and A. M. Rahmani, “A multi-objective approach for energy-efficient and reliable dynamic VM consolidation in cloud data centers,” *Engineering science and technology, an International Journal*, vol. 26, p. 100995, 2022.

[14] A. Ponraj, “Optimistic virtual machine placement in cloud data centers using queuing approach,” *Computing Systems*, vol. 93, pp. 338–344, 2019.

[15] M. Terra-Neves, I. Lynce, and V. Manquinho, “Virtual machine consolidation using constraint-based multiobjective optimization,” *Journal of Heuristics*, vol. 25, no. 3, pp. 339–375, 2019.

[16] Z. Li, X. Yu, L. Yu, S. Guo, and V. Chang, “Energy-efficient and quality-aware VM consolidation method,” *Future Generation Computer Systems*, vol. 102, pp. 789–809, 2020.

[17] S. Jin, X. Qie, W. Zhao, W. Yue, and Y. Takahashi, “A clustered virtual machine allocation strategy based on a sleepmode with wake-up threshold in a cloud environment,” *Annals of Operations Research*, vol. 293, no. 1, pp. 193–212, 2020.

[18] H. Mora, F. J. Mora Gimeno, M. T. Signes-Pont, and B. Volckaert, “Multilayer architecture model for mobile cloud computing paradigm,” *Complexity*, vol. 2019, Article ID 3951495, 13 pages, 2019.

[19] S. S. Gill, S. Tuli, A. N. Toosi et al., “ThermoSim: deep learning based framework for modeling and simulation of thermal-aware resource management for cloud computing environments,” *Journal of Systems and Software*, vol. 166, p. 110596, 2020.

[20] C. Ghribi, *Energy Efficient Resource Allocation in Cloud Computing Environments, PhD Thesis*, Institut National des Télécommunications, 2014.

[21] M. Gao, T. Xie, S. S. Du, and L. F. Yang, “A provably efficient algorithm for linear Markov decision process with low switching cost,” 2021, http://arxiv.org/abs/2101.00494.

[22] S. K. Panda and P. K. Jana, “An energy-efficient task scheduling algorithm for heterogeneous cloud computing systems,” *Cluster Computing*, vol. 22, no. 2, pp. 509–527, 2019.