IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 13, 2018, accepted April 22, 2018, date of publication April 25, 2018, date of current version May 16, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2830183

A Multi Criteria-Based Approach for Virtual Machines Consolidation to Save Electrical Power in Cloud Data Centers

MOHAMMED AMOON
Department of Computer Science, Community College, King Saud University, Riyadh 28095-11437, Saudi Arabia
Computer Science and Engineering Department, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt
e-mail: mamoon@ksu.edu.sa

This work was supported by King Saud University, Deanship of Scientific Research, Community College Research Unit.

ABSTRACT Consolidation of virtual machines is used to reduce the power consumed in cloud computing systems. In consolidation, some virtual machines are migrated from some source servers to other destination servers and source servers are turned off. Most current consolidation approaches depend on the utilization of servers to determine both source and destination servers. In this paper, a consolidation approach that depends on multiple criteria is proposed and evaluated. The approach has one algorithm for determining source servers and another algorithm for determining destination servers. Simulations experiments show relevant improvements over utilization-based approach in terms of throughput, power consumption, monetary cost, and scalability by 21%, 12%, 24%, and 37%, respectively.

INDEX TERMS Consolidation, throughput, virtual machines, data center, power consumption.

I. INTRODUCTION

Cloud is a highly dynamic environment and then allocation of virtual machines to servers of data centers cannot ensure the guarantee of the service quality expected by customers or the benefits or profit expected by providers. Balancing the amount of power consumption and the performance of a cloud represents a critical issue. Inefficient employment of resources is one of the most significant sources of ever-increasing power consumption in these systems. This means that more resources are used in executing applications than the sufficient number of resources and then more power is consumed [1].

In cloud computing environments, greater power saving can be achieved by using consolidation, Dynamic Voltage Frequency Scaling (DVFS), or efficient scheduling of customer applications. In consolidation, the cloud tries to use fewer servers or physical machines in order to execute applications without breaking the Service Level Agreement (SLA) [2], [3]. So, the cloud selects some servers to be turned off and their virtual machines or applications are migrated or transferred to other servers where they can complete execution. In DVFS, the computing elements can change their voltages and frequencies according to the allocated applications [4]. In addition, scheduling of applications to virtual machines and scheduling of virtual machines to servers can play important role in reducing energy consumed in cloud environments [5], [6]. Most existing research works are consolidation based or scheduling based because the property of DVFS is not always available in cloud computing systems due to the high cost. In addition, the performance of the cloud can be degraded when the frequency of the computing elements is reduced.

The research work developed in this paper is based on consolidation technique. In this work, the main focus is to propose a multi criteria-based consolidation approach that depends on more than one criterion when selecting both source and destination servers. The main contributions introduced in this paper are:

1) Proposing an algorithm for determining the source servers to be turned off and the list of virtual machines to be migrated.
2) Proposing an algorithm for determining the destination servers on where the migrated virtual machines will be executed.
3) Integrate both of the proposed algorithms into an approach in order to obtain superior performance in terms of the most important performance criteria for both provider and customer.

The next sections of this paper are organized as follows: Section 2 briefly explains consolidation-based related work, Section 3 presents a brief description of the problem statement, Section 4 elaborates in details the proposed approach, Results are presented and discussed in Section 5 and Section 6 provides the conclusion and some suggestions for the future.

II. RELATED WORK

Many researchers have extensively studied managing power consumption in cloud computing systems and multiple methods were proposed. The proposed methods are classified as methods based on software techniques such as efficient scheduling and consolidation and methods based on hardware techniques such as DVFS. The research work in this paper is based on consolidation technique. The concept behind this technique is to utilize as minimum servers as possible in order to reduce power consumed by data centers. This can be done by migrating some selected virtual machines from some source servers to other destination servers and switching off the source servers and then energy can be saved [2], [3].

Multiple consolidation methods and approaches were developed in the context of reducing power consumption.

Murtazaev and Oh [7] have proposed a consolidation mechanism to reduce the number of utilized servers and virtual machine migration. However, their mechanism considers only utilization of memory and CPU when selecting the virtual machined to be migrated in the data centers.

Rao and Thilagam [8] have proposed a consolidation mechanism, called residual resource defragmentation that uses Sercon algorithm for reducing the number of active servers. The mechanism depends on exchanging virtual machines among servers to reduce defragmentation.

Beloglazov and Buyya [9] proposed a consolidation algorithm, which ranks the virtual machines according to CPU utilization and allocates each virtual machine to servers by turns. Then, for each server, it calculates the amount of increased energy and selects server with minimum increased energy as the target server. Beloglazov *et al.* [10] proposed a method for reducing the migration of virtual machines depending on utilisation thresholds.

Han *et al.* [11] proposed a resource utilization aware consolidation algorithm to improve resource utilization and power consumption. The algorithm considers both utilization state and number of active recourses and it consolidate underutilized servers to minimize migration of virtual machines. Liao *et al.* [12] proposed a consolidation approach that sorts servers according to their CPU utilization, performs migration of virtual machines from the underutilized servers and turns down some inactive servers to save energy.

Lee and Zomaya [13] proposed two consolidation methods that consider active and idle servers. The main base of the their methods is to maximize utilization of servers and minimize energy consumption.

Similarly, Horri *et al.* [14] have developed a scheduling approach that depends on the utilization of the employed servers. Arianyan *et al.* [15] proposed a consolidation based strategy depending on CPU, memory and bandwidth of servers to determine the servers with heavy loads. In [16], they extend their work by combining the consolidation and the DVFs methods in a fuzzy based approach with the aim to minimize the conflict between the two methods. Frincu [17] developed a mechanism to avoid improper decisions of virtual machine migrations that have considerable influence on the cloud profit. Esfandiarpoor *et al.* [18] proposed consolidation methods that depend on switching down some other devices such as cooling and network devices to save energy. Beloglazov and Buyya [19] have developed a dynamic mechanism for consolidation of virtual machines. They have proposed a set of algorithms for defining overloaded and underloaded servers, determining virtual machines to be migrated and locating the new servers for the virtual machines to be migrated.

Reviewing previously developed works reveals that most proposed consolidation methods consider the utilization of servers as the main criterion for selecting source and destination servers. However, in addition to utilization, there is a set of other effective criteria that can be considered when determining these servers such as failure probability and power consumption rate. In this paper, failure probability and power consumption rate of servers are also considered for determining both source and destination servers.

Researches believed that Utilization is a good indicator for the amount of power consumed by servers in data centers, so they used it for determining source and destination servers for the migration of virtual machines. In most cases, utilization is not the only appropriate indicator for determining source and destination servers and this represents the motivation of research work in this paper. For example, if we have two servers S1 and S2 and the utilization of each of them is 60% and 10%, respectively. The failure probability of S1 is 0.7 and the failure probability of S2 is 0.1. The power consumption rates of S1 and S2 are 40KWh and 20KWh, respectively. Which server will be more suitable in this case S1 or S2? Utilization based consolidation method will select S2 as the source server to be turned off after migrating virtual machines to S1. However, S1 is prone to fail more than S2 and in case of failure of S1 more power will be consumed to alleviate the effect of this failure. In addition, the power consumption rate of S1 is higher and worse than the power consumption rate of S2. So, S1 will consume more power than S2. Thus, depending only on utilization of servers is not a good choice.

III. PROBLEM STATEMENT

Customers submit their applications with the service requirements to the cloud. The cloud assigns the applications to the most appropriate virtual machines according to the service requirements.

Generally, the service requirements comprise response time and financial cost of performing their applications in the data center. Response time of an application is the interval time between submitting the application and obtaining results

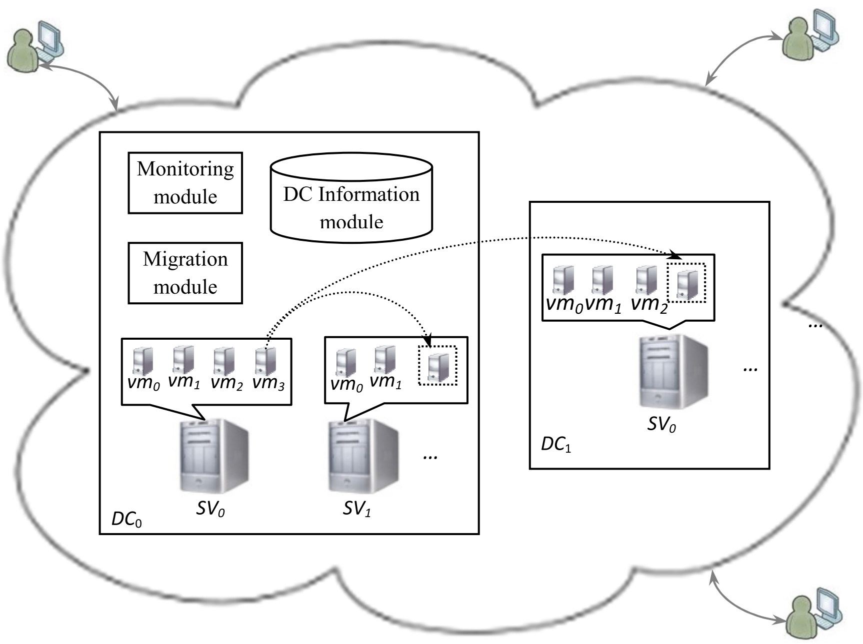


FIGURE 1. The cloud architecture.

of its execution. It is defined as:

$$\tau_i = \tau_{ic} + \tau_{ie} + \tau_{iw}, \tag{1}$$

where τ_{ic} is sum of the time required to transfer the application i from the customer to the assigned virtual machine and the time required to transfer execution results to the customer, τ_{ie} is the time required to process the application i in the data center and τ_{iw}, is the expected waiting time incurred by the application [20].

The financial cost represents the price the customer will pay for the cloud provider for service performed. Most providers offer pay-per-use service in which customers pay only for resources used by their applications. For example, Google chargers customers according to number of CPU cycles used [21].

Virtual machines are allocated to servers or physical machines in the cloud. Each server can host multiple virtual machines. During the execution of these applications, servers that host the allocated virtual machines could expend unnecessary electrical power and then produce more financial cost which is not acceptable by both customers and providers.

Consolidation through migration of virtual machines is a popular technique used in order to reduce the power consumption through reducing the number of employed servers to the most possible number. So, some virtual machines are transferred or migrated to other servers and the same number of virtual machines can be applied with fewer servers. Then, the released servers are switched off and thus the amount power consumption is reduced.

The decision of migration is taken according to the migration strategy applied by the cloud. Selecting the migration strategy is a major challenge. In other words, why should the migration of a virtual machine be performed? Most existing consolidation techniques consider the utilization of servers as the criterion when decision of migration is taken. However, in addition to utilization, there are also some other important factors that can affect the performance of the data center and thus they should be also considered. These factors can include failure probability, response time, power consumption rate, monetary cost rate, load balancing, etc.

IV. THE PROPOSED CONSOLIDATION APPROACH

In this section, the proposed multi-criteria based consolidation approach is presented. The approach has two algorithms in the context of saving electrical power. The first algorithm determines the set of source servers to be turned off and group of virtual machines to be migrated. However, the second algorithm determines the destination servers for these virtual machines.

Determining the strategy used in migration is one of the cloud provider responsibilities because it has the full awareness about the infrastructure of the cloud. The provider of the cloud expects that the migration process of virtual machines contributes in improving the performance of the cloud. Therefore, the main rule of migration is to transfer virtual machines from servers of poor performance to others of better performance. In order to efficiently migrate virtual machines or applications, a search must be conducted through the cloud to find the destination server to which the virtual machine will be migrated and where the performance will be improved. The criterion on which this search is based represents the major challenge in the migration process.

In this paper, the architecture shown in Figure 1 is considered. Each cloud contains one or more data centers and

Input: S is the set of turned on servers in data center d,
V_s is the set of VMs run on server s in data center d,
$U(x), x = 1, 2, \ldots, n$, represents the utilizations of S,
$U_{lth}(x), x = 1, 2, \ldots, n$, represents the lower threshold values of utilization of S,
$F(x), x = 1, 2, \ldots, n$, represents the failure probabilities of S,
$F_{th}(x), x = 1, 2, \ldots, n$, represents the threshold values of the failure probability of S,
$P(x), x = 1, 2, \ldots, n$, represents the average power consumption of S,
$P_{th}(x), x = 1, 2, \ldots, n$, represents the threshold values of average power consumption of S,
$T(x), x = 1, 2, \ldots, n$, represents the response time of S,
$T_{th}(x), x = 1, 2, \ldots, n$, represents the threshold values of the response time of S,

Output: $SS \subset S$, is the set of source servers,
M, is the set of VMs to be migrated from source servers of d,

```
For each s_i ∈ S do
   If (U(i) < U_lth(i) || F(i) > F_th(i) || P(i) > P_th(i) || T(i) > T_th(i))
      {
      Add s_i to SS; //add the server to the set of source servers
      For each v_ij ∈ V_i do //add the set of VMs to the migration list;
         Add v_ij to M;
      EndFor
      }
   EndIf
   Shut down s_i;
   Remove s_i from S;
EndFor
```

FIGURE 2. The SSS algorithm.

Input: S_x is the set of servers, received from DC Information module, that can host the virtual machine x,
M is the set of VMs to be migrated,
$U_{uth}(x), x = 1, 2, \ldots, n$, represents the upper threshold values of utilization of S,

Output: RD_x is the list of recommended destination servers for virtual machine x,
D_x is the set of destination servers (one server for each virtual machine x)

```
Sort M according to the required priority for each virtual machine.
For each x ∈ M do //list of recommended servers
   For each s_i ∈ S_x do
      If (U(i) < U_uth(i) || F(i) < F_th(i) || P(i) < P_th(i))
         Add s_i to RD_x
      EndIf
   EndFor
EndFor
For each x ∈ M do // list of destination servers
   For each s_i ∈ RD_x do
      Compute SF
   EndFor
   Allocate x to the server with the largest value of SF in RD_x
   Add this server to the D_x
EndFor
```

FIGURE 3. The DSS algorithm.

each data center has a set of servers or physical machines on which virtual machines are constructed and implemented to run applications of customers. Also, each data center contains a set of modules to monitor and control the operation of its resources. These modules include Monitoring module, Migration module and DC Information module. The main duty of the Monitoring module is to observe the performance of all virtual machines in the data center and to collect the up to date status data of all resources in the data center. This data is then sent to the DC Information module in order to update the related records. The DC Information module has a data base that contains up to date status data needed about data center servers.

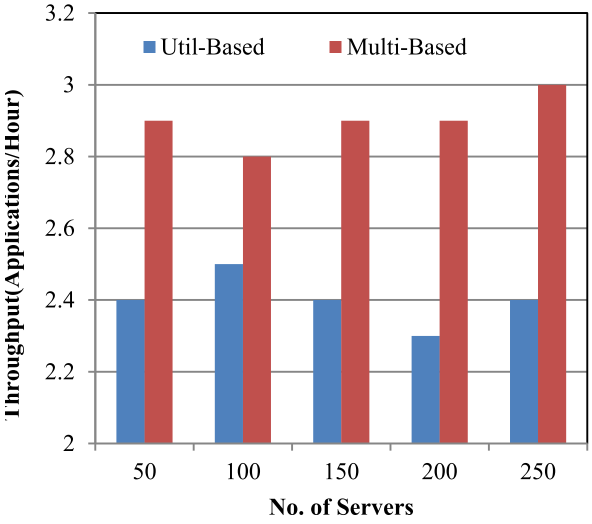
According to the performance requirement, the Monitoring module prepares a set of source servers to be turned off and a migration list that contains virtual machines to be migrated. Then, it delivers this list to the Migration module. The Monitoring module implements the logic of the Source Server Selection (SSS) algorithm shown in Figure 2 in order to build the migration list. The SSS algorithm selects virtual machines from servers that have performance under threshold values of failure probability, utilization and power consumption.

For each virtual machine, the Migration module asks the DC Information module for a set of servers that can host the virtual machine according to the customer's requirements. The DC Information module supplies the migration module with a list of servers for each virtual machine in the migration list. Based on this servers list, the Migration module determines the list of recommended servers for each virtual machine. The set of servers in the recommended list should satisfy the threshold values of power consumption rate, failure probability and utilization. Then, the most suitable destination server is selected from the recommended list according to the policy defined by the Destination Server Selection (DSS) algorithm shown in Figure 3. The selected server is added to the list of destination servers.

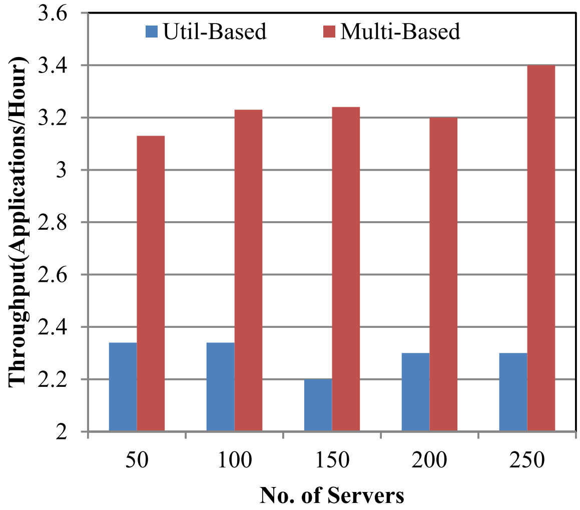
In order to achieve that, the DSS algorithm computes the value of Selection Factor (SF) for each server in the set of recommended servers and uses this value to select the most suitable destination server for each virtual machine in the migration list. The Selection Factor (SF) of a server r is defined by:

$$SF_r = \sum_{k=0}^{m} A_{rk} W_{rk}, \tag{2}$$

where A_{rk} represents a criterion partially used to determine the value of the Selection Factor of server r. The criteria used may include server utilization, failure probability, response time, power consumption rate, monetary cost rate, load balancing, etc. In accordance to their needs, the provider and the customer determine the value of each related criterion. For example, the provider determines the values of failure probability and utilization criteria, while the customer determines the values of response time and monetary cost criteria. The value W_{rk} is the corresponding weight for each criterion. It is noted that, the sum value of all weights should equal 1.



(a)



(b)

FIGURE 4. The comparison of throughput. (a) 1000 VMs. (b) 5000 VMs.

For each virtual machine, the algorithm selects the server with the largest value of Selection Factor from the list of the recommended destination servers and then dispatches the virtual machine to the selected destination server.

V. RESULTS

The use of simulation tools enables research investigators to develop and assess their proposed methods without the need to access actual clouds. CloudSim is a standout among simulation tools created for that purpose. In spite of the fact it helps in simulating the methods of the main aspects and issues related to cloud computing it has no support for simulating consolidation based methods. Along this, extra modules ought to be created in order to enable and empower the simulation of consolidation-based methods and approaches.

A. SIMULATION SETUP

The performance of the approach proposed in this paper is evaluated using experimental simulation. For this purpose, a cloud of 10 data centers is considered and each data center can implement about 1000 virtual machines with processing speed ranges from 20 to 100 GHz. The assumed power consumption rate ranges from 10 to 100 Kilo-Watt/Hour and the assumed speed of the network link is 1 Gb/s. The response times required for executing customers' applications ranges from 10 to 1000 Hours. For all the following experiments, the number of servers in each data center ranges from 50 to 300 and the experiments is performed with 1000 and 5000 virtual machines distributed over the data centers.

Failures of servers are simulated using the Poisson probability distribution as in [22]. The Poisson distribution assumes the failure probability of a server S in two disjoint time intervals is:

$$F_S(X) = \frac{e^{-\mu}\mu^x}{x!}$$
$$0 \leq F_S(X) \leq 1 \quad and\ x = 0, 1, 2, \ldots, n, \qquad (3)$$

where $X(x_0, x_1, x_2, \ldots, x_n)$ is the number of failures in a certain time interval and μ is the mean number of failures of this interval for a server S.

In order to assess the proposed approach, it is compared versus the utilization-based consolidation approach introduced in [6]. The comparison metrics include throughput, monetary cost, amount of power consumption and scalability. The displayed results are acquired by utilizing different arrangements of servers and virtual machines of cloud data centers.

B. THROUGHPUT

In distributed and high performance computing systems, throughput is an important metric used to evaluate performance. It measures the amount of applications the system can process over a given period of time [23] and it represents an indication for the capacity of the data center. In clouds, the throughput of a data center d is defined as:

$$\text{Thr}(d) = \frac{m}{t_m}, \qquad (4)$$

where m represents the number of applications assigned to d and t_m is the required time to complete m applications.

In Figure 4(a and b), the throughput comparison between the proposed multi criteria-based approach and the utilization-based approach is depicted. The x-axis acts as the number of servers of a data center and the y-axis acts as the throughput resulted. The throughput of both approaches grows with the increment of the number of servers. However, it is obvious that the proposed approach provides better throughput than the utilization-based approach. This is because the proposed approach considers the failure probability of servers when migration decision is taken. On the other hand, the utilization-based approach does not consider the failure probability of servers. So, some servers can fail and then the execution of all their assigned virtual machines should be restarted again from scratch on other servers. Therefore, a significant amount of time is added to the execution time of these virtual machines. Even fault tolerance techniques are applied, the time will increase because the copies of the virtual machines will be executed on slower servers than the original failed server. As long as the execution time increases the throughput decreases.

C. POWER CONSUMPTION

The consumed power by a data center represents the amount of electrical energy used during its running over the unit time. For data centers, Power Usage Effectiveness (PUE) is the most well known measure for the power efficiency. It was launched by the Association of Green Grid [24] and it is defined as:

$$\mathrm{PUE} = \frac{\text{Total Power}}{\text{IT Power}}, \tag{5}$$

where the Total Power represents electrical power consumed by all facilities of the data center including IT resources of the data center, generators, cooling units, etc. The IT Power represents the electrical power consumed by the IT related resources such as servers, memory units and network infrastructure devices.

For a data center d, the amount of consumed IT Power for a virtual machine v is denoted by PIT_{dv} and it contains processing power, data communication power and storage power. It can be defined as follows:

$$PIT_{dv} = PP_{dv} + PC_{dv} + PS_{dv}, \tag{6}$$

where PP_{dv} represents the amount of processing power, PC_{dv} represents the amount of data communication power and PS_{dv} represents the amount of storage power consumed by virtual machine v.

The Total Power consumed by a virtual machine v in a data center d is denoted by P_{dv} and it is defined as:

$$P_{dv} = PIT_{dv} + PNIT_{dv}, \tag{7}$$

where $PNIT_{dv}$ represents the amount of power consumed via all other non IT resources.

Figure 5(a and b) shows the comparison of power consumption between the proposed multi criteria-based approach and the utilization-based approach. The x-axis acts as the number of servers of a data center and the y-axis acts as the amount of power consumed via these servers. It is obvious that the proposed approach has less power consumption than the utilization-based approach. This is due to that the proposed approach considers the amount of power consumed via servers when determining both source and destination servers. Where, the servers of high power consumption will be recommended as source servers. This consideration gives the advantage of little power consumption to the multi criteria-based approach.

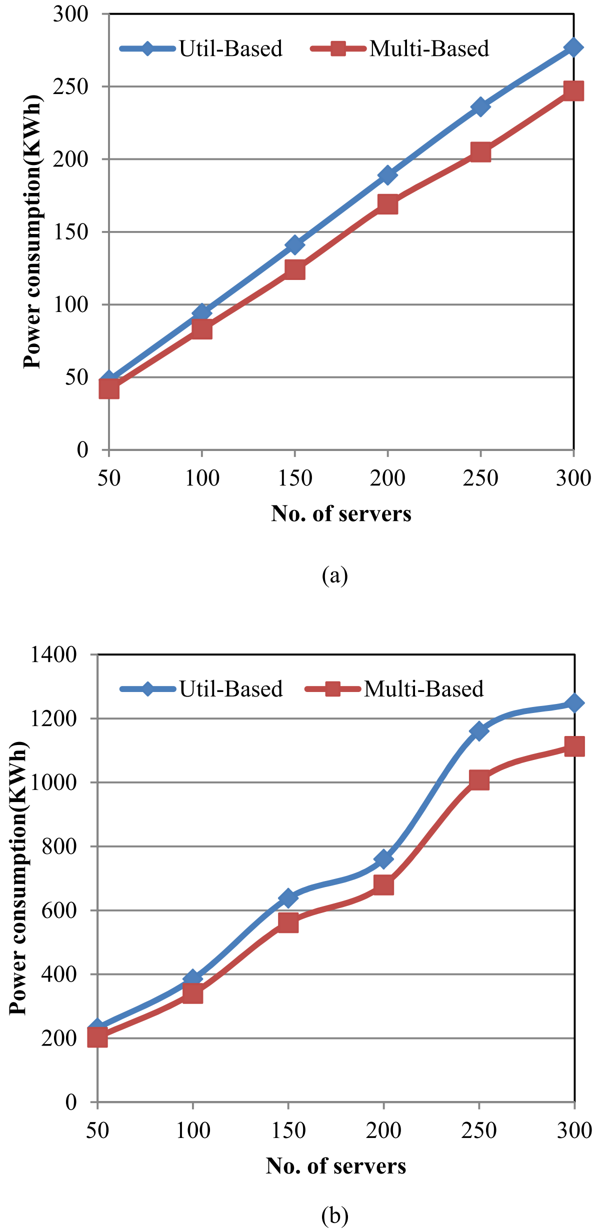


FIGURE 5. The comparison of power consumption rate. (a) 1000 VMs. (b) 5000 VMs.

D. MONETARY COST

The amount of monetary cost paid for cloud services represents a major issue for cloud customers. It is favorable for customers to reduce this amount to the most possible extent without violating the quality required for their services.

In Figure 6, the comparison of monetary cost of executing virtual machines on servers of data center between the proposed multi criteria-based approach and the utilization-based approach. The x-axis acts as the number of virtual machines or customers' applications and the y-axis acts as the amount of monetary cost. When compared with the utilization-based approach, the multi-criteria based approach shows superiority. This is because the multi criteria-based

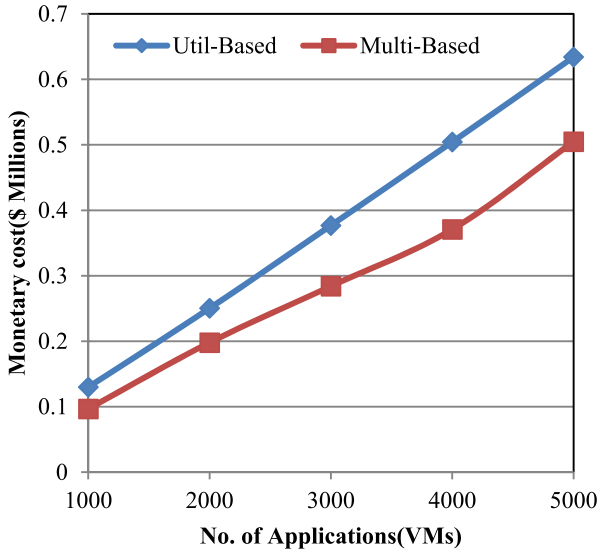


FIGURE 6. The comparison of Monetary cost.

TABLE 1. Performance improvement.

Metrics	Percentage of improvement
Throughput	21%
Power consumption	12%
Monetary cost	24%
Scalability	37%

approach considers failure probability and power consumption when determining both source and destination servers. Considering failure probability decreases failures and therefore no extra cost will be added due to re-execution of virtual machines from scratch. Also, considering power consumption decreases the monetary cost due to using servers of high power consumption.

E. SCALABILITY

Scalability of a cloud refers to the extent to which the cloud can carry out the customers' applications without violation of customers' requirements. The better the scalability of a cloud, the more applications it can carry out at the same time. So, scalability determines the ability of the system to deal with a large number of customers' application simultaneously.

Figure 7(a and b) depicts the comparison of scalability improvement between the proposed multi criteria-based approach and the utilization-based approach. The x-axis of the figure acts as the number of customers' applications and the y-axis acts as the percentage of scalability improvement. In general, both approaches improve the scalability of the cloud. However, it is clear that the percentage of scalability improvement of the proposed approach is better than that of the utilization-based approach. This is because considering multiple criteria leads to increasing the chance for carrying out applications without violating the requirements of customers.

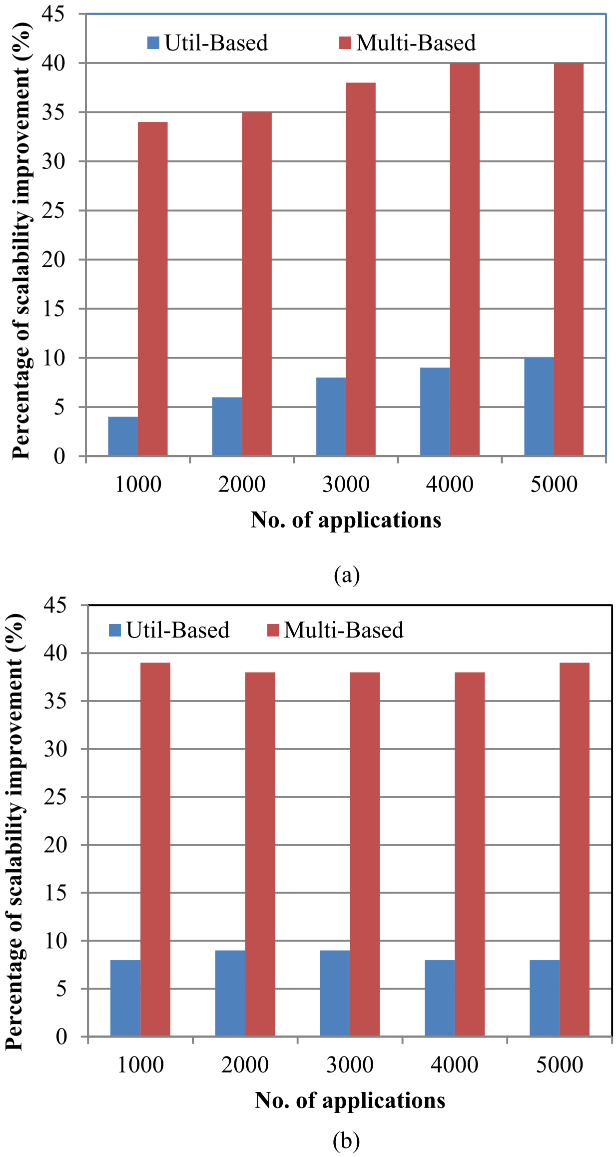


FIGURE 7. The comparison of scalability. (a) 200 servers. (a) 400 servers.

VI. CONCLUSIONS

Consolidation is a technique used to reduce power consumption of data centers through turning off some servers. In this paper, a multi criteria-based consolidation approach is presented and evaluated. The approach has an algorithm that determines servers to be turned off and another algorithm that determines destination servers that will execute virtual machines of turned off servers. The performance of the proposed approach is compared against a recent utilization-based approach. The proposed multi criteria-based approach has superior performance than the utilization-based approach in terms of throughput, power consumption rate, monetary cost and scalability as shown in Table 1.

REFERENCES

[1] W. Zhu, Y. Zhuang, and L. Zhang, "A three-dimensional virtual resource scheduling method for energy saving in cloud computing," *Future Generat. Comput. Syst.*, vol. 69, pp. 66–74, Apr. 2017.

[2] S. S. Manvi and G. K. Shyam, "Resource management for infrastructure as a service (IaaS) in cloud computing: A survey," *J. Netw. Comput. Appl.*, vol. 41, pp. 424–440, May 2014.

[3] Y. Gao, H. Guan, Z. Qi, T. Song, F. Huan, and L. Liu, "Service level agreement based energy-efficient resource management in cloud data centers," *Comput. Elect. Eng.*, vol. 40, pp. 1621–1633, Jul. 2013.

[4] C.-M. Wu, R.-S. Chang, and H.-Y. Chan, "A green energy-efficient scheduling algorithm using the DVFS technique for cloud datacenters," *Future Generat. Comput. Syst.*, vol. 37, pp. 141–147, Jul. 2014.

[5] F. Fernandes, D. Beserr, E. D. Moreno, B. Schulz, and R. C. G. Pinto, "A virtual machine scheduler based on CPU and I/O-bound features for energy-aware in high performance computing clouds," *Comput. Elect. Eng.*, vol. 56, pp. 854–870, Nov. 2016.

[6] D. Buia, Y. Yoon, E.-N. Huh, S. Jun, and S. Lee, "Energy efficiency for cloud computing system based on predictive optimization," *J. Parallel Distrib. Comput.*, vol. 102, pp. 103–114, Apr. 2017.

[7] A. Murtazaev and S. Oh, "Sercon: Server consolidation algorithm using live migration of virtual machines for green computing," *IETE Tech. Rev.*, vol. 28, no. 3, pp. 212–231, 2011.

[8] K. S. Rao and P. S. Thilagam, "Heuristics based server consolidation with residual resource defragmentation in cloud data centers," *Future Generat. Comput. Syst.*, vol. 50, pp. 87–98, Sep. 2015.

[9] A. Beloglazov and R. Buyya, "Energy efficient resource management in virtualized cloud data centers," in *Proc. 10th IEEE/ACM Int. Conf. Cluster, Cloud Grid Comput.*, Melbourne, VIC, Australia, 2010, pp. 826–831.

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

[11] G. Han, W. Que, G. Jia, and W. Zhang, "Resource-utilization-aware energy efficient server consolidation algorithm for green computing in IIOT," *J. Netw. Comput. Appl.*, vol. 103, pp. 205–214, Feb. 2018.

[12] X. Liao, L. Hu, and H. Jin, "Energy optimization schemes in cluster with virtual machines," *Cluster Comput.*, vol. 13, pp. 113–126, Jun. 2010.

[13] Y. C. Lee and A. Y. Zomaya, "Energy efficient utilization of resources in cloud computing systems," *J. Supercomput.*, vol. 60, pp. 268–280, May 2012.

[14] A. Horri, M. Mozafari, and G. Dastghaibyfard, "Novel resource allocation algorithms to performance and energy efficiency in cloud computing," *J. Supercomput.*, vol. 69, pp. 1445–1461, Sep. 2014.

[15] E. Arianyan, H. Taheri, and S. Sharifian, "Novel heuristics for consolidation of virtual machines in cloud data centers using multi-criteria resource management solutions," *J. Supercomput.*, vol. 72, pp. 688–717, Feb. 2016.

[16] E. Arianyan, H. Taheri, and V. Khoshdel, "Novel fuzzy multi objective DVFS-aware consolidation heuristics for energy and SLA efficient resource management in cloud data centers," *J. Netw. Comput. Appl.*, vol. 78, pp. 43–61, Jan. 2017.

[17] M. E. Frîncu, "Scheduling highly available applications on cloud environments," *Future Generat. Comput. Syst.*, vol. 32, pp. 138–153, Mar. 2014.

[18] S. Esfandiarpoor, A. Pahlavan, and M. Goudarzi, "Structure-aware online virtual machine consolidation for datacenter energy improvement in cloud computing," *Comput. Elect. Eng.*, vol. 42, pp. 74–89, Feb. 2015.

[19] 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 Comput., Pract. Exper.*, vol. 24, no. 13, pp. 1397–1420, Sep. 2012.

[20] M. Amoon and T. Tobely, "A green energy-efficient scheduler for cloud data centers," in *Cluster Computing*. New York, NY, USA: Springer, 2018. [Online]. Available: https://doi.org/10.1007/s10586-018-2028-z

[21] S. Su, J. Li, Q. Huang, X. Huang, K. Shuang, and J. Wang, "Cost-efficient task scheduling for executing large programs in the cloud," *Parallel Comput.*, vol. 39, pp. 177–188, Apr./May 2013.

[22] M. Amoon, "Adaptive framework for reliable cloud computing environment," *IEEE Access*, vol. 4, pp. 9469–9478, 2016.

[23] M. Amoon, "A fault-tolerant scheduling system for computational grids," *Comput. Elect. Eng.*, vol. 38, no. 2, pp. 399–412, 2012.

[24] (Dec. 2017). [Online]. Available: https://www.thegreengrid.org/en/resources/library-and-tools/20-PUE%3A-A-Comprehensive-Examination-of-the-Metric

MOHAMMED AMOON received the B.Sc. degree in electronic engineering and the M.Sc. and Ph.D. degrees in computer science and engineering from Menoufia University, Shibin El Kom, Egypt, in 1996, 2001, and 2006, respectively. He was appointed as an Assistant Professor with Menoufia University in 2006, where he is currently an Associate Professor with the Computer Sciences and Engineering Department. He is also an Associate Professor with the Computer Science Department, King Saud University. His research interests include agent-based systems, fault tolerance, green computing, distributed computing, grid computing, cloud computing, and fog computing.