An Energy-efficient Virtual Machine Placement Algorithm based Service Level Agreement in Cloud Computing Environments

Arash Mazidi
Dep. Computer Engineering
Islamic Azad University of Aliabad Katool Branch
Aliabad Katool, Iran

Elham Damghanijazi
Dep. Computer Engineering
Islamic Azad University of Aliabad Katool Branch
Aliabad Katool, Iran

Sajad Tofighy
Dep. Computer Engineering
Shiraz University
Shiraz, Iran

ABSTRACT
The cloud computing has given services to the users throughout the world during recent years. The cloud computing services have been founded according to ‘As-Pay-You-Go’ model and some leading enterprises give these services. The giving these cloud-computing services has been developed every day and these requirements necessitate for more infrastructures and Internet providers (IPs). The nodes of data centers consume a lot of energy in cloud structure and disseminate noticeable amount of carbon dioxide into the environment. We define a framework and structure for cloud environment of efficient energy in the present paper. We examine the present problems and challenges based on this structure and then present and model management algorithms and source allocation in cloud computing environment in order to manage energy in addition to considering Service Level Agreement. The proposed algorithm has been implemented by cloudsim simulator where the obtained results from simulation of real-time data indicate that the proposed method is superior to previous techniques in terms of energy consumption and observance of Service Level Agreement. Similarly, number of live migration of virtual machines and quantity of transferred data has been improved.

Keywords

1. INTRODUCTION
Today, cloud computing has been known and established as pioneer process in proposing computational services throughout the world [1]. The majority of the given cloud-computing services are presented within the frameworks of Infrastructure as Service, Platform as Service, and Software as Service [2]. Virtualization is one of the important features of cloud-computing. The virtualization lets several virtual machines to operate as a single physical machine at the same time. One can refer to improvement in efficiency of sources, lower costs, and easier management of servers as one of the advantages of this technology. As the users pay cost against using services they expect to receive appropriate Quality of Services from the provider [3].

Therefore, if the provider could not deliver the favorable quality of services to the user, he should pay fine to user for this violation. The criteria for analysis on quality of services depend on general performance, response time, and delay time of the given service. The Service Level Agreement is a part of agreements between service- providers and users where time of delivery of service and/ or efficiency of service are specifically assumed as some of these cornerstones [4].

The cloud-computing is also deemed very important in terms of energy consumption. The data centers are composed of thousands of computational nodes which, in turn, consume a lot of electrical energy. Presently, two topics of energy consumption and rate of violation from Service Level Agreement have been turned into one of the foremost challenges regarding cloud-computing. The computers of these centers are responsible for the paramount part of energy consumption at a cloud-computing center since these centers constitute thousands of computational nodes. According to a report published by IDC, the rate of consuming power by nodes and cooling systems has been totally estimated 30 million USD throughout the world; as the energy consumption shown has noticeably increased for 10 years period between years 2000-2010 [5].

The dynamic management and composition of virtual machines is one of the conspicuous techniques for reducing violation from Service Level Agreement and amount of energy consumption [6]. In This method, we use live migration of virtual machines from one host to another as a powerful tool to prevent from violation against Service Level Agreement and reducing energy consumption. If the rate of efficiency of a host is low (Under Load) all of virtual machines on this host can migrate to the other hosts to prevent from waste of energy consumption and also make the aforesaid host to turn off mode. In contrast, due to variability of workloads on the hosts, a certain host may reach to higher load mode (Over Load) in which one or more virtual machines located on that host should be selected to be transferred onto other hosts in order to balance workload in this host for observance of Service Level Agreement. Awareness of Service Level Agreement means that the algorithm specially addresses reducing violation from Service Level Agreement and it can reduce remarkably rate of violation from Service Level Agreement by decrease in consuming power [6].

This paper is mainly focused on proposing an algorithm of efficient energy by considering Service Level Agreement. In other words, it is tried to prevent violation from Service Level Agreement as possible to reduce operational costs.

We will explore the formerly conducted studies in details and propose the new technique. Then we will analyze the results compared to the previous methods and conclude at the end.
2. RELATED WORKS
Most of the presented studies on energy management at virtualized data centers are based on live migration of virtual machines and minimization of quantity of hosts by the aid of turn-off process and/ or changing them into lower energy consumption mode in idle hosts. The energy consumption is reduced by means of these techniques and on the other hand efficiency of resources increases as well. Given this point that accurate minimization of the consuming power is crucially important since doing of this Improperly may lead errors take place in Service Level Agreement (SLA Violation) and payment of fine and finally to increase sum of costs. Many algorithms are proposed for solving this problem as heuristic algorithms and others [7].

The first study was conducted on management of power inside virtualized aiding systems by Nathuji and Schawn. The local manager is responsible for taking management techniques of virtual machines at any host. The overall manager may select the suitable host for migration of virtual machines by observation of efficiency in several various hosts as well. This technique is not efficient because the hosts of Operating Systems are not aware of energy [8].

Kusic et al. have assumed resource management in cloud-computing environment as an ordinal optimization problem and solved it by the aid of LLC. This activity aimed to minimize rate of energy consumption and percent of violation from Service Level Agreement. The author utilized Kalman filter [10] to approximate future status. The main problem of this technique lies in complexity of controller so that it needs to 30min time for 15 nodes and therefore this technique will be inefficient for data centers at large scale [9].

Verma et al. have formulated dynamic placement of programs as a packing problem including boxes with variable sizes. The live migration is used for transferring virtual machines to new locations. The Service Level Agreement was not considered in this paper [11].

Beloglazov and Buyya divided this problem into four parts including 1) finding of overload hosts, 2) selection of virtual machines out of overload host, 3) finding of underload host, and 4) selection of destination for virtual machines (of overload and underload hosts). They also proposed some exploratory techniques for these sections. The more prominent techniques proposed in this study are as follows. Local Regression Algorithm is used for identifying overload hosts and Minimum Migration Time technique is employed for selection of virtual machines of overload host [12]. Similarly, Power Aware Best Fit Decreasing algorithm is an energy-aware packing technique and used for selection of destination to virtual machines. Finally, simple method was used for identifying underload hosts. The best composition of the given exploratory techniques was also adaption of these above-said methods. The resource management is conducted only based on efficiency of processor of virtual machines and hosts and it does not address the status of other parameters in virtual machines of hosts such as efficiency of CPU. Therefore, the main problem of this method lies in restriction of hypothesis of the given problem only to efficiency of processor and this has caused to reduce maneuvering power tangibly for resource management to the authors [13].

Horri et al. have assumed problem of optimization of energy consumption as a 2D packing problem by considering rate of efficiency of RAM memory along with efficiency of CPU memory and given some methods for solution. The given strategies for phase of identifying of underload hosts, placement of migrant virtual machines, and allocation of source to the given migrant virtual machines showed that it has led to improvement in consumed power by the hosts along with reduced rate of violation from Service Level Agreement [14].

Rahmanian et al. have reformulated problem of optimization of the consuming power by the hosts so that to convert it into a cost optimization problem. As a result, by considering operational costs in cloud data centers including costs of consuming power and fine arisen of violation from Service Level Agreement, the author indicated that both of these criteria might lead to increase of costs for providers. Accordingly, some solutions were proposed to maximize profit for the providers by taking operational costs as aware-costs. The suggested solution by the authors has led to improvement in rate of profit resulting for the cloud provider [15].

Liu et al. in their paper, evolutionary computing is applied to VMP to minimize the number of active physical servers, so as to schedule underutilized servers to save energy. Inspired by the promising performance of the ant colony system (ACS) algorithm for combinatorial problems, an ACS-based approach is developed to achieve the VMP goal [16].

Chen et al. have focused on the energy efficiency and thermal stability issues of the cloud systems. A Cross Entropy based VM Placement (CEVP) algorithm is proposed to simultaneously minimize the energy cost, total thermal cost and the number of hot spots in the data center. Simulation results indicate that the proposed CEVP algorithm can (1) achieve energy savings of 26.2% on average, (2) efficiently reduce the temperature cost by up to 6.8% and (3) significantly decrease the total number of the hot spots by 60.1% on average in the cloud systems, by comparing to the Ant Colony System-based algorithm [17].

3. PROPOSED APPROACH
The current algorithm used for managing virtual machines comprises of four steps: 1) finding of overload hosts, 2) selection of virtual machines in overload host, 3) finding of underload hosts, and 4) selection of destination for the virtual machines.

In all of the given previous methods, only current status of efficiency and history of efficiency status belonged to virtual machines and hosts have been utilized for making decisions and managing data center by the aid of four aforesaid steps since it has been assumed in the former studies that the rate of efficiency of CPU might depend on level of efficiency in processor. As a result, only efficiency of processor is employed as an important feature for making decisions. However, by analysis on output data from real data centers we concluded that the rate of dependency among efficiency of processor and CPU in various virtual machines might totally differ from the assumptions in the previous studies and unlike what it considered, the rate of quantitative dependency of these two parameters is totally ignorable. To prove this claim, we analyzed rate of dependency among efficiency of processor and CPU in a great number of the rendered virtual machines by the aid of regression for 24h and in 9 different days and finally came to this result that the rate of dependency of CPU on efficiency of processor has been averagely less than 13% in them at all times of running of virtual machines. For this reason, we will adapt dynamic resource management rather than use of efficiency of processor and use the efficiency of CPU in virtual machines as well as along with the process as an important feature. Likewise, we will take
help from an exploratory technique to predict trend and for selection of virtual machines with ascending trend to transfer and for reducing placement of hosts at overload mode in order to select virtual machines in overload hosts for transmission.

### 3.1 Problem Formulation

Equation (1), shows the objective function of the problem. As seen, this is a multi-objective function while Service Agreement Violation (SLAV) reduction increase power consumption and vice versa. Equation (2) calculates total power consumption of all hosts during runtime. Let \( n \) denote the number of time slots, \( H_n \) denotes the number available hosts in time slot \( t \), \( V_{n}^{j} \) is the number of VMs of host \( j \) in time slot \( t \) and \( VU_{n}^{j} \) denotes the CPU utilisation of VM \( k \) in time slot \( t \) (see Equation (4)). Equation (3), calculates average SLAV of hosts during runtime. Let \( H_i \) is the host number \( i \) and \( SLAV(H_i) \) denotes the SLAV rate of Host \( i \) (see equation (5)) [18].

\[
\text{Minimize} \ (\varphi, \rho) \quad (1)
\]

\[
\varphi = \sum_{i=1}^{n} \sum_{j=1}^{H_{n}} \text{Power} \left( \sum_{k=1}^{V_{n}^{j}} VU_{n}^{j} \right) \quad (2)
\]

\[
\rho = \frac{1}{H_n} \sum_{i=1}^{H_n} \text{SLAV}(H_i) \quad (3)
\]

Subject to:

\[
t = \{1, 2, \ldots, n\} \quad (4)
\]

\[
0 \leq VU_{n}^{j} \leq 100 \quad (5)
\]

#### 3.2 The Proposed Algorithm for VMs Selection in Overloaded Hosts

The minimum migration time was the previous algorithm presented in this section where in fact the main idea in this algorithm was the migration of virtual machines at the minimum migration time and this policy would lead to reducing reduction in total consuming energy. However, the main objective in above-said algorithm is to prevent from overhead status of the given host that was not addressed and it was sufficed with idea of selection of virtual machines with the minimum migration time for transmission. The defect of this technique lies in this fact that it is unlikely to select virtual machines for migration with descending trend in efficiency while other virtual machines were available inside the hosts acted as the main factor to make the host as overloaded and they might have strongly ascending and or upward trend. Nonetheless, due to policy for current choice, they have not been selected for migration. This event will lead to this trend that although the host has currently exited from overload mode, there will be very high likelihood for placement of this host in overload status in the near future and rather than increasing possible placement of host at overload mode, this event may increase total energy consumption due to inefficient excessive migrations as well and consequently rate of violation from Service Level Agreement as well.

Therefore, to this end, we propose an algorithm for this step to cover this fault. The given algorithm in this section initially calculates trend of each of virtual machines inside overload host to select migration of virtual machines and it will select only those virtual machines in which ascending trend is higher than the threshold level as candidates for transmission. The aforesaid threshold level is also computed and considered according to lifetime of virtual machine. Now based on this algorithm, only those virtual machines with ascending trend may be selected as subjects to placement of host at overload status for transmission out of these candidates one can select the virtual machines with the criterion of the minimum migration time for transmission. Thus, we repeat operation for transmission of virtual machine according this criterion to the extent the condition for escape from overload mode satisfies.

#### 3.3 The proposed Algorithm for Destination Selection of VMs Migrations

The former technique proposed for the algorithm is the best energy-aware adjustment mode where due to this nature it is a single dimensional packing algorithm may not be used for this task since according to nature of our problem; the single dimensional packing algorithm could not be utilized by adding parameter of efficiency of CPU. This event has occurred because this problem is two-dimensional; therefore two-dimensional bin packing algorithm should be used for solution. There are several two-dimensional bin packing algorithms for solving this problem and they be used for solving the given problem. The best algorithm for this space is one that increases rates efficiency in two dimensions equally in any host. Thus with respect to this issue, we have chosen vector bin packing algorithm as a criterion for bin-packing.

The reason for selection of this technique is related to growing efficiency in CPU and processor (both dimensions of coordinates) proportional and aligned with each other inside the hosts. This technique tries to select a host as destination of virtual machine where according to Equation (6), efficiency of host in both dimensions includes the minimum difference from the main axis in two-dimensional plane of coordinates after host placement.

\[
\begin{align*}
\text{Arg}_{\text{Max}} \ & \ CPVM \times (1 - CPVM_{\text{Host}}) + \text{RAM}_{\text{VM}} \times (1 - \text{RAM}_{\text{Host}}) \\
\end{align*}
\]

In addition to vector bin packing technique, we employ another bin-packing method called Norm Based Greedy technique due to its proportional nature to this problem. In this method, Equation (7) is used as criterion to select destination for any virtual machine between various hosts. In this formula, the host is obtained with the minimum value resulting from the given formula.

\[
\begin{align*}
\text{Arg}_{\text{Min}} \ & \ ((1 - CPVM_{\text{Host}}) + (1 - \text{RAM}_{\text{Host}}) \times (\text{CPVM} + \text{RAM}_{\text{VM}})^2)
\end{align*}
\]

### 4. IMPLEMENTATION AND RESULTS

In this section, we examine simulator architecture, properties of virtual machines, and workload data.

#### 4.1 Simulator

The cloudsim simulator is utilized as a tool for implementation, testing, and review of algorithms for examination and analysis of results with reproduction potential. Thus, using simulators e.g. cloudsim may be vitally important for analysis on efficiency of algorithms under various configurations. Therefore, we employ cloudsim simulator as an intermediate factor for review and assessment of study in the current investigation.

#### 4.2 Workload data

We use data from CoMon project (Park, 2006) for 9 days in this study. These data include efficiency of processor and CPU for every 5min belonged to more than one thousand virtual machines placed inside more than 8000 hosts.
4.3 Evaluations
We will discuss on mean assessment of efficiency in the given algorithm in this section. We have run the algorithm for several times in order to achieve the results in these algorithms and finally we have discussed the findings by computation of mean value for several running processes in this study. The variables of energy consumption, rate of violation from Service Level Agreement, and number of migration of virtual machines as the mean findings of assessment of simulation are given respectively in the table.

In the Table 1, the proposed algorithm has been compared by basic techniques i.e. LR/MMT, LR/MU, LR/MC, and LR/RS. The conducted analyses and assessments signify that the rates of energy consumption and violation from Service Level Agreement have been improved and additionally numbers of live migrations of virtual machines and volume of transferred data were increased. The policies given for dynamic norm-based process of cloud sources may act noticeably better than static policies. Adaption of the suggested framework and given algorithm for analysis phase to compose virtual machines dynamically may propose better results than technique used in the previous policies and this means that using former records of hosts has acted very efficiently in placement of virtual machines.

Table 1. The simulation results

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Energy, kWh</th>
<th>SLAV*10^-5</th>
<th>VM Migration</th>
<th>Shut down</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>146.52</td>
<td>3.26</td>
<td>15492.2</td>
<td>3834</td>
</tr>
<tr>
<td>LR/MMT</td>
<td>161.87</td>
<td>4.974</td>
<td>28174.7</td>
<td>4593.85</td>
</tr>
<tr>
<td>LR/MU</td>
<td>173.06</td>
<td>6.28</td>
<td>29419.2</td>
<td>5426.7</td>
</tr>
<tr>
<td>LR/RS</td>
<td>119.27</td>
<td>7.89</td>
<td>23692</td>
<td>4061</td>
</tr>
<tr>
<td>LR/MC</td>
<td>148.51</td>
<td>7.60</td>
<td>23931.2</td>
<td>4116.6</td>
</tr>
</tbody>
</table>

Table 2 shows the rate of improvement of the present solution compared to other aforesaid techniques based on criteria for assessment of quality of efficiency (rate of violation from Service Level Agreement). As it shown in this table, due to migration of proposed solution, efficiency rate has been reduced. Decrease in total number of migration (PDM) in this strategy was the reason for reducing rate of efficiency due to migration of given strategy (As it characterized in Table 4-4). Moreover, the period of placement of hosts at overload statue (SLATAH) in the given solution has been tangibly reduced in comparison with other assessed strategies. As a result, total rate of violations from Service Level Agreement (SLAV), as the product of two primary assessment criteria, has been also highly reduced in the suggested solution.

Table 2. The comparison of the efficiency and violation from Service Level Agreement in various algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>PDM</th>
<th>SLATAH</th>
<th>SLAV*10^-5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>0.00815</td>
<td>0.004</td>
<td>3.26</td>
</tr>
<tr>
<td>LR/MMT</td>
<td>0.0082</td>
<td>0.006</td>
<td>4.974</td>
</tr>
<tr>
<td>LR/MU</td>
<td>0.00897</td>
<td>0.007</td>
<td>6.28</td>
</tr>
<tr>
<td>LR/RS</td>
<td>0.0087</td>
<td>0.009</td>
<td>7.89</td>
</tr>
<tr>
<td>LR/MC</td>
<td>0.0095</td>
<td>0.008</td>
<td>7.60</td>
</tr>
</tbody>
</table>

Table 3 displays rate of improvement in criteria of assessment of quality of efficiency of the proposed strategy in percent in comparison with other evaluated solutions. As it identified, there was noticeable improvement in all of criteria for the proposed strategy compared to other solutions. For example, it is visible that the proposed solution includes the lesser SLA violation (58.68%) than LR/RD solution.

Table 3. The comparison of improvement percent in quality of efficiency and violation from Service Level Agreement for the proposed strategy compared to the other solutions

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>PDM</th>
<th>SLATAH</th>
<th>SLAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR/MMT</td>
<td>0.6%</td>
<td>33.33%</td>
<td>34.45%</td>
</tr>
<tr>
<td>LR/MU</td>
<td>9.14%</td>
<td>42.85%</td>
<td>48.08%</td>
</tr>
<tr>
<td>LR/RS</td>
<td>6.32%</td>
<td>55.55%</td>
<td>58.68%</td>
</tr>
<tr>
<td>LR/MC</td>
<td>14.21%</td>
<td>50%</td>
<td>57.10%</td>
</tr>
</tbody>
</table>

Figure 1 indicates that rate of energy consumption has been remarkably reduced. In fact, the proposed algorithm causes the hosts at normal status not to be selected as overload and underload candidates and as a result number of active and turn-off hosts as well as quantity of turn-on processes for the host with number of migrations to be decreased so this is led to reduction in consuming total power.

The proposed algorithm causes the underload hosts to be converted into passive mode. Therefore number of active hosts is reduced and following to reducing number of active hosts, rate of static energy consumption also decreases thus total efficiency is increased in the system.

![Fig 1: Diagram for comparison of power consumption by various methods at different days of workload](image-url)
prevents the hosts in destination of virtual machines to be placed at overload mode.

In addition, there is the least possibility for rising workload in virtual machines of overload hosts in the future due to the conducted approximation and this leads to prevention of destination hosts from overload in the virtual machines and total number of migrations is reduced.

Figure 3 shows that the Service Level Agreement has been noticeably improved by the proposed framework and algorithm. As a result, with respect to reduction in number of placement of hosts at overload and underload modes, rate of violation from Service Level Agreement has decreased. Consequently, as you see in Figure 4, number of turn-off and turn-on of cloud resources has been remarkably reduced.

In addition to reduced rate of energy consumption caused by lower migration costs dependent on that factor, number of subsequent turn-off and turn-on in cloud hosts also affects in rate of violation from Service Level Agreement. Number of migration of virtual machines has been reduced due to selection of overload machines based on the given algorithms according to estimator model of driving mean since this practice prevents from multiple migrations of virtual machines among various hosts.

In addition, there is the least possibility for rising workload in virtual machines of overload hosts in the future due to the conducted approximation and this leads to prevention of destination hosts from overload in the virtual machines and total number of migrations is reduced.

5. CONCLUSION
New method was proposed in this paper for resource management in cloud computing space with respect to energy consumption and Service Level Agreement. The given findings from simulation of real data indicate that the presented technique is superior to the previous techniques in terms of energy consumption and observance of Service Level Agreement. The analyses and assessments signify that the rates of energy consumption and violation from Service Level Agreement have been improved and moreover numbers of live migrations of virtual machines and volume of trained data were increased. The given policies act noticeably better than static policies for dynamic norm based process of cloud resources. Use of the suggested frameworks and proposed algorithm may propose better results than the previous policies for phase of analysis to compose virtual machines dynamically and this means use of previous background of the hosts has acted very efficiently in placement of virtual machines. Then we intend to use the relevant data to the network and risk of making decisions as well.

6. REFERENCES


