Smart Computing Review

Smart Virtual Machine Placement Using Learning Automata to Reduce Power Consumption in Cloud Data Centers

Hossein Ghiasi[1] and Mostafa Ghobaei Arani[2*]

[1]Department of Computer Engineering, Islamic Azad University of Mahallat / Mahallat , Iran / Ho3in.ghia30@gmail.com

[2]Department of Computer Engineering, Islamic Azad University of Parand / Tehran, Iran / mostafaghobaei@piau.ac.ir

*Corresponding Author: Mostafa Ghobaei Arani

Received October 12, 2015; Revised November 15, 2015; Accepted December 19, 2015; Published December 31, 2015

Abstract: Today, cloud computing is one of the most challenging research topics in the field of information technology. It is so important for computer researchers that it was included on a list of top ten technologies in the world. Data centers include reservoirs where processing power can meet the needs of many users' computing. The popularity and acceptance of cloud computing has increased the number of these centers in recent years. One of the challenging issues in cloud computing environments is high energy consumption in data centers, which has been ignored in the corporate competition to develop data centers. High energy consumption by data centers leads to increased costs, as well as CO2 emissions. Researchers are now struggling to find an effective approach to decrease energy consumption in data centers. In recent years, many attempts have been made to reduce the power consumption of data centers, and many approaches have been proposed to reduce power consumption, such as hardware and software approaches and approaches using virtualization technology. In fact, placement of a virtual machine (VM) means finding a suitable physical place for the VM. The placement goal can either maximize the usage of available resources or it can save power by being able to shut down some servers. In this paper, we present an approach based on a best-fit decreasing (BFD) algorithm, which uses learning automata to reach a compromise between decreasing energy consumption and violating service level agreements.

Keywords: Learning Automata, Minimum Correlation Coefficient, Reduction of Consumed Power, Reduction of Service Level Agreement Violations, Virtual Machine Consolidation, Virtual Machine Placement

DOI: 10.6029/smartcr.2015.06.005

Introduction

Data centers are one of the main components of cloud computing. In recent years, information technology **center**(ITC) service providers like IBM, Microsoft, Google and other similar organizations have developed data centers to increase performance in computing.

The focus on the efficiency of the data centers led to ignoring energy consumption by these centers. In 2006, the cost of energy consumption for data centers was US$405 billion, with an increasing trend (according to the predictions) that will be double that for 2011 [1]. As well, the consumed data center energy accounted for 1.5% of the total energy consumption of the USA [2]. The data centers of Facebook, Google and YouTube, respectively, were 10.52%, 7.74% and 3.27% of the total energy consumption of the IT departments [3]. The average energy consumption of data centers is equivalent to 25,000 households [4]. In addition to operational costs, high energy consumption increases temperature and reduces the reliability and longevity of hardware resources. Due to the considerable production of carbon dioxide (CO_2), greenhouse gas emissions and environmental issues have also been addressed [5], such that using power management techniques seems essential [6]. Fortunately, efforts to reduce power consumption have been conducted, in part, for CPU power [7], for power management and for power management of storage and networks [8].

Different sources account for the energy consumed by a server in a data center. Each server has resources, such as the CPU, RAM, network equipment and storage resources. As seen in Figure 1, a comparison with other system resources shows that CPUs consumed more energy than other resources [9].

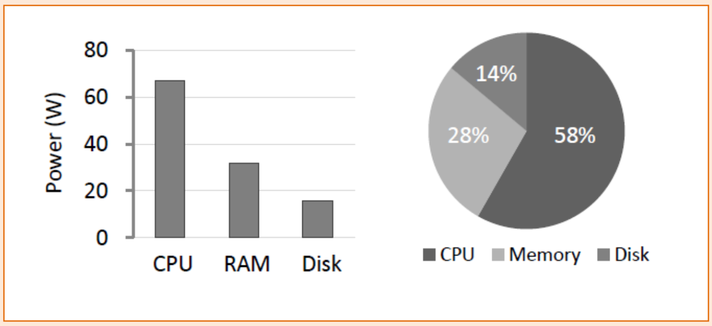

Figure 1. The power consumption of a server's resources [9]

Idle servers are one of the most important factors wasting energy in cloud data centers that have low utilization of energy-consuming hardware resources, because servers do not have power fit. This means that on a light load or even at idle, servers consume a large amount of energy, compared to the times of maximum utilization of their resources [10]. For example, during six months, Google's researchers studied 5000 physical machines and found more times when the physical machines ran at between 10 to 40 per cent efficiency, and the machines rarely achieved efficiency close to 100% .

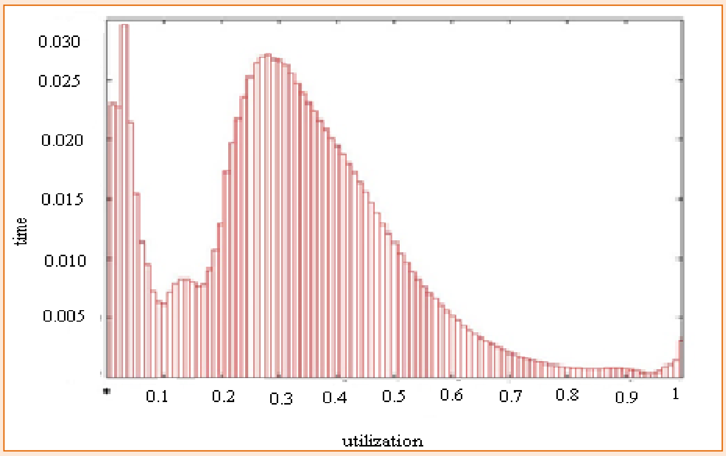

Figure 2. The utilization of a sample of 5000 Google servers

Virtualization techniques are solutions that lead to reduction in energy consumption by data centers. Virtualization technology enables multiple virtual machines (VMs) located on a single physical machine (PM) or server, so that it is possible to reduce the amount of hardware resources and improve productivity [12-13]. One of the other benefits of virtualization is relatively easy VM transfer by a technique called *live migration*, which easily transfers a virtual machine to another virtual machine. Live migration transfers still-operating VMs from one physical machine to another PM without interrupting the operation of the VM [14]. The PM, according to VM requests and the needs of available resources (such as the CPU, RAM, bandwidth, etc.), runs the VMs assigned to it. The number of VMs may be limited on a server. As a result, some PMs have low efficiency of resources. A new method developed to address this problem is called immigration. All VMs on PMs with low productivity are transferred to another PM by this technique, and the idle PMs go to sleep or shut down, so the number of active PMs is decreased. That process is called server consolidation. But there is a question here. Which virtual machine should be transferred to which physical machine? The issue referred to as "the placement of the virtual machine" is known. In fact, placement of a virtual machine means finding a suitable physical machine for the VM [15].

Placement of virtual machines is done at two different times. The first situation occurs when a new VM is created. This virtual machine has to be allocated to a host that can provide the resources needed. That is called initial allocation. The second situation is when, according to the terms of agreement, a specific performance level is necessary, and that virtual machines is migrated to another physical machine, called the Target, which is scheduled to host the virtual machine. The virtual machine placement process is one of finding a suitable host for the virtual machine. In other words, the virtual machine placement is mapping virtual machines to physical machines [15].

For a small number of virtual and physical machines, it might be possible for a human operator to manage the placement of virtual machines. However, as the number of VMs and PMs increases, the situation becomes unmanageable, and automation becomes necessary. Even in an automated calculation of placement plans, the number of possible mappings to be evaluated for a given set of virtual and physical machines is [16] by equation 1:

$$(\textit{Number of physical hosts})^{(\textit{Number of virtual machines})} \qquad (1)$$

Such a large solution space makes it almost impossible for any brute force VM–PM mapping algorithm to provide exact placement plans in practical, acceptable times. Hence, intelligent placement heuristics is needed to narrow down the search to a solution that obtains near-optimal placement plans [17].

Our problem statement can be briefly described as follows:

- *M* physical machines are available, and their resource capacities given, such as memory CPU power and network bandwidth dimensions.
- There are *N* virtual machines to be placed. The requirements of these virtual machines are given, such as memory, CPU power and network bandwidth.
- We have to find a mapping between VMs and PMs that satisfies the VMs' resource requirements while minimizing the number of physical machines used.

When finding such a mapping, we have to take care that the total resource requirements of the VMs placed on a PM must not exceed the PM's capacity, because it reduces the quality of service, resulting in violation of service level agreements.

The article continues in Section 2 with related work, in Section 3, with a discussion of the proposed method, and in Section 4, with an evaluation of the efficiency of the proposed method.

Related work

Various research has been performed to solve the problem of placement of virtual machines using heuristic algorithms, e.g., first fit (FF), best fit (BF), based on constraint programming, random number programming, evolutionary algorithms, etc.

Beloglazov and Buyya [18] used an integrated virtual machines technique to solve the energy problem. It detects overloaded and under-loaded hosts using upper and lower utilization thresholds to reach a compromise between energy reduction and violating a service level agreement (SLA). If the host has lower efficiency than a lower utilization threshold, its virtual machines should be transferred to another host. Then, the host with zero productivity will shut down. They were applying a best-fit decreasing (BFD) approach to solve placement issues, and their algorithm is known as modified best-fit decreasing (MBFD). They considered various policies to detect additional overloaded hosts, and also reviewed various policies to select different virtual machines for stages of integration of virtual machines, and compared them. Results of their research showed that the local regression–minimum migration time (LR-MMT) policy has the best outcome with respect to energy. They considered other parameters, such as multiple consumed power and percentage of violation of service level agreements, to compare policies.

Hieu and Soonwook [19] performed the placement of a virtual machine using a BF greedy algorithm and hierarchical clustering in two stages. In the first stage, virtual machines are located on the same physical machine with high traffic. In the next stage, according to clustering resources, placement of virtual machines is performed. Then, minimum cuts of a binary tree are used to optimize the network issues to reduce the number of active nodes and the overall energy consumption.

Panigrahy et al. [20] used BFD for placement of virtual machines. They try to decrease network traffic along with energy consumption. The relationships between virtual machines has been modeled and, by weighting the relations, a graph is drawn according to the virtual machines and their weights; then, the weighted graph is converted to a hierarchical model, considering the weight of its crests, which reviews network traffic for placement of the virtual machines. This considers low productivity of the physical server for VM migration, and a VM will be migrated without reduction in efficacy.

Xu et al. [21] selected a virtual machine that has the lowest value from dot multiplication of two vectors for resources and capacity. To achieve suitable productivity required a VM that needs a faster CPU and less memory to be placed on a physical machine with a slower CPU and higher memory capacity. This approach may be an incorrect selection, because it does not measure the length of vectors.

Xiaoli and Zhanghui [22] addressed the problem of live migration by considering limits and dot vector multiplication of resources and the available capacity of the host. There is an event evoker in the suggested model, which receives feedback of information from resources, including the start of the VM, the size of the VM, VM migration, and so on. Then, it sends a map and a plan of placement to the supervisor and addresses the size and number of migrating VMs to prevent sequential and non-beneficial migrations. The approach controls consumed energy by considering limits, and uses a changed FF algorithm for bin packing.

Goodarzi and Pedram [23] addressed placement of VMs in a California university using several copies of the VM and dynamic programming, as well as local search. Then, using these copies, they optimized a compromise between consumed energy and productivity. Several copies are usually used to increase credibility. The main VM is responsible for offering the requested service from the service provider. The other VMs should be kept inactive till they are needed, so that inactive VMs cause an increase in availability and productivity, and decrease consumed energy. They practiced dynamic programming and linear search for placement of VMs to solve the problem of bin packing.

Kord and Haghighi [24] offered an approach for VM placement where the objective was compromise between a decrease in energy and a violation of the SLA. They used a minimum correlation coefficient to locate the virtual machine. They used a fuzzy analytic hierarchy process to reach a compromise between decreasing energy and violation of the SLA. Also, they offered a model for focused management. The offered algorithm works in a manner where a VM that wants to select a host selects a host with minimum consumed energy and a minimum correlation coefficient.

Singh and Hemalatha [25] offered a hierarchy algorithm using a honey bee–comb algorithm to decrease energy consumption according to placement, which determined the best place for the VM according to the CPU, using VM clustering. Also, it decreased delay of transmission, and they tried to have minimum live migration in VMs.

Hasanul et al. [26] focused on the balance of server resource efficacy using an ant colony paradigm. Their objective was to decrease power and reduce resource waste. They operated it by vector algebra. They modelled the host's capacity, the VMs demands and resource efficacy as vectors, and they considered a matrix for VM placement where the elements show the PM to host the VM. Also, they worked on resources waste by using vectors, and found the best mapping among possible mappings by considering an ant colony.

Feller et al. [27] addressed integration of workload using the ant colony paradigm. They solved the problem of multi-dimensional packing by optimizing the ant colony. In their proposed algorithm, each of the ants receives all packings (for example, VMs) and all bins (for example, physical machines) and starts to select a host according a probable decision-making rule, which describes optimization of work for the ant. It updates the amount of pheromone, and puts the second virtual machine on the host. The rule guides the ants, based on the information about current pheromone concentration, to choose the most promising items. So, according to the amount of pheromone and information, the ant will choose the pack with the higher probability. The random nature of the algorithm lets the ants find many potential solutions to explore. Finally, when all the ants' solutions have been found, the amount of pheromone is determined (vaporizing or increasing) according to the pheromone updating rule.

Jiang et al. [28] formulated energy issues using the placement of the VM in a cloud architecture with multi-dimensional resources. They offered a solution for reduction of power consumption in data centers using three local search algorithms and a genetic algorithm solution, but it did not have enough sustainability. They considered other computing resources, included amount of RAM and number of disks, in addition to the CPU. They considered hosts homogeneous, and virtual machines as heterogeneous resources. They addressed the problem as a multi-dimensional packing issue.

Rasouli et al. [29] presented a new method based on learning automata to replace virtual machines in data centers dynamically to reduce energy consumption. The main policy of this approach is reduction of live migration, and switching an idle node to inactive or sleep mode. Each physical node in a data center cloud is an automaton, where its learning includes acceptance or rejection of the virtual machine from the source node to the destination. An automaton has two actions, where only one is active at a time. They considered four states based on the efficiency of the processor, which includes idle, middle, active and overloaded statuses. Hosts with zero CPU efficiency will be considered a shut-down host. Transition from one state to another is equivalent to switching from one node to another after accepting the virtual machine, the same as a Markov chain. Performance of the host was divided into host–sender, receptor–host and neutral, for proper management of migration of virtual machines. Every host can remain in its previous state or locate itself in various states, according to its current situation. Coefficients of reward or penalty for learning automata are values that are determined based on original and destination node status. The destination node will be used in the next practice, according to the environment response, so that the newly created allocation will be closer to the optimal schedule. Repetition will be performed so many times in order to find a suitable placement, so there will be no other changes. They just focused on energy, so SLA violation is presented in our method, more than the other approaches.

Proposed approach

Because our approach is according to learning automata, and because placement policy requires a multiple correlation coefficient, in the following, we address the introduction of the multiple correlation coefficient and learning automata.

■ Learning Automata

A learning automata [30-34] is an abstract model that can perform a limited number of actions. The selected action is assessed by random environment, and a response is given to the learning automata. The automaton uses the response and selects its action for the next stage. Figure 3 shows the relationship between learning automata and environment.

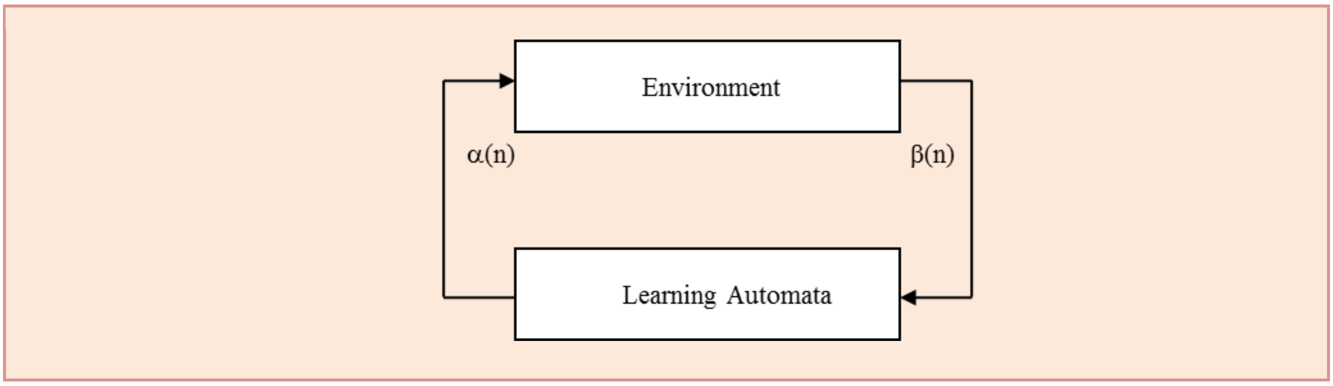

Figure 3. Relationship between environment and automaton

In this paper, we use learning automata with various structures. Learning automata with various structures is shown by five pieces: $\mathbf{LA} \equiv \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \mathbf{p}, \mathbf{T}, \mathbf{c}\}$.

$\alpha \equiv \{\alpha_1, \alpha_2, \alpha_3 \cdots, \alpha_r\}$ shows a set of actions in learning automata
$\beta \equiv \{\beta_1, \beta_2, \beta_3 \cdots, \beta_m\}$ shows an input set of the automata
$p \equiv \{p_1, p_2, p_3, \dots . p_n\}$ is vectors for the possibility of choosing an action-learning algorithm
$T \equiv p(n+1) = T[\alpha(n), \beta(n), p(n)\,]$
$c \equiv \{c_1, c_2, c_3, \cdots, c_r\}$ is the possibility of finding any action

If the learning automata chooses the action α_i in stage i and receives a favorable response from the environment, $P_i(n)$ will increase and the other probabilities will decrease. From a favorable response, $P_i(n)$ will decrease, and the other probabilities will increase. Changes are made to the sum of $P_i(n)$ such that it always remains constant and equal to 1.

i. Favorable response will be made by equation 2.

$$p_i(n+1) = p_i(n) + \alpha[1 - p_i(n)] \qquad (2)$$
$$p_j(n+1) = (1-b)p_j(n) \ \forall j \ j \neq i$$

ii. A favorable response will be made by equation 3.

$$p_i(n+1) = (1-b)p_i(n) \qquad (3)$$
$$p_j(n+1) = \frac{b}{r-1} + (1-b)p_j(n) \ \forall j \ j \neq i$$

In the above equation, a shows reward, and b shows penalty.

■ Multiple correlation coefficient

A multiple correlation coefficient is used for assessment of the prediction quality of the dependent variable in multiple regression analysis. Its value depends upon predicted values and the real value of the dependent variable. The value of a dependent variable Y results from the set of explaining variables: x1, x2,…,x3. The relation of observation i from the Y variable results from a variable P, as is shown in equation 4.

$$y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_p x_{np} + e_i \quad (4)$$

in which Y_i is the i value of the dependent variable. P is the number of predictions. b_j is the j coefficient, and X_{ij} j= 0 , … , p, which is the i value for the j prediction. e_i is a violation in the observed value for the i case.

Suppose that X is a matrix of N*(P+1), in which the first column is always 1, and includes data from the independent variable and the Y vector, which is a vector of 1*N from the observed real value of the dependent variable. These matrixes are shown in equations 5 and 6.

$$X = \begin{matrix} 1 & \dots & x_{1p} \\ \vdots & \dots & \vdots \\ 1 & \dots & x_{np} \end{matrix} \qquad (5)$$

$$y = \begin{matrix} y_1 \\ \vdots \\ y_n \end{matrix} \qquad (6)$$

The vector of the predicted value of the dependent variable y is shown as $\hat{y}$, which results from equation 7:

$$\hat{y} = Xb \ b = (XX^T)^{-1}X^t y \qquad (7)$$

X^T is the transposed X matrix, the quality of prediction is assessed by calculating the multiple correlation coefficient, which is shown with $R_{2y,1\ldots,p}$ (equation 8), which measures the degrees of two linear variables, and its value is between zero and 1. Zero indicates there is no linear relationship between the two. So, if there is a linear relationship with a positive slope, we will have a value of 1.

$$R^2_{y.1\ldots.p} = \left[\frac{\mathrm{cov}(y_i,\hat{y})}{\sqrt{\mathrm{var}(y_i)\mathrm{var}(\hat{y})}}\right] \qquad (8)$$

Here COV shows covariance, and Var shows variance. The above equation will be shown as equation 9, too.

$$R^2_{y.1\ldots.p} = \left[\frac{\sum_1^n (y_i - m_y)^2 (\hat{y_i} - m_{\hat{y}})^2}{\sum_1^n (y_i - m_y)^2 \sum_1^n (\hat{y}_i - m_{\hat{y}})^2}\right] \qquad (9)$$

In this equation, m_y and $m_{\hat{y}}$ show the mean of y and $\hat{y}$ to calculate the correlation between the VM and VMs on the selected physical machines. Here, matrix X includes the productivity of the VMs on the host. Columns of the matrix include M productivity of the history of a VM that is operating on the selected host, and the vector of Y includes M productivity of the history of the selected VM.

■ Proposed algorithm

Here, the number of automaton actions equals the number of hosts. The automaton is allowed to have only one active action at any one time. In other words, the automaton selects only one host from the list of candidate hosts at any one time. The vector of automaton probability is an array called P [I]. Its primary value equals 1/(Number of Host). The amount of P[I] shows the probability of selecting a host where the ID equals i.

The automaton selects from the candidates list one host that has enough resources for the VM and that will not be overloaded after dedicating resources to the VM. Then, the list of virtual machines on the host is obtained. The amount of power after being selected to host the current virtual machine is estimated, in order to estimate the energy efficiency of the linear relationship between the CPU and the energy.

The amount of energy is compared to the energy that will be consumed by the other hosts after accepting the current VM. This process is done for correlation among the current VM and the other hosts' VMs. If the current host has a minimum amount of energy and correlation, a reward is dedicated to the automaton.

These stages are repeated for the number of hosts, and finally, the host with the higher probability will be selected as the destination host.

Because successive rewards increase the possibility of selecting some hosts while there have become to host Overload.. To restore the hosts who reach overload status, a reduced probability of their choice, compared to the rest of the differences, is divided between the other hosts.

Performance Evaluation

We consider four kinds of VM for simulation, which are shown in Table 1. Also, the types of PMs are shown in Table 2, and the number of VMs on the various days is shown in Table 3.

Table 1. Types of VMs

Type	Size	BW	RAM	PE	MIPS
1	2.5 GB	100Mbit/s	870	1	2500
2	2.5 GB	100Mbit	1740	1	2000
3	2.5 GB	100Mbit	1740	1	1000
4	2.5 GB	100Mbit	613	1	500

Table 2. Types of PMs

Type	Storage	BW	RAM	PE	MIPS
1	1GB	1Gbit/s	4096	2	1860
2	1GB	1Gbit/s	4096	2	2660

Table 3. The number of VMs on the various days

20110420	20110412	20110411	20110409	20110403	20110325	20110322	20110309	20110306	20110303	Date
1033	1054	1233	1358	1463	1078	1516	1061	898	1052	**The Number Of VMs**

We used a thresholds.(THR) selection policy to detect overloaded hosts, and we determined a 0.8 error confidence level. Also, to select the virtual machine, we used an MMt policy. The results of the simulation were compared with dynamic voltage and frequency scaling (DVFS)and the method offered by Beloglazov called MBFD. Figure 4 and Figure 5 show the results of the simulation.

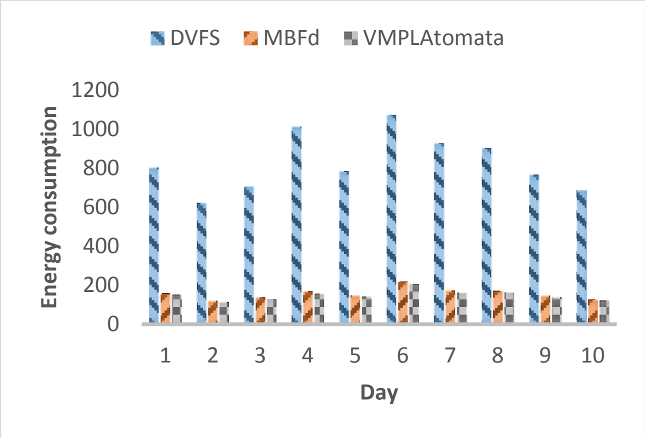

Figure 4. Energy consumption in the three approaches

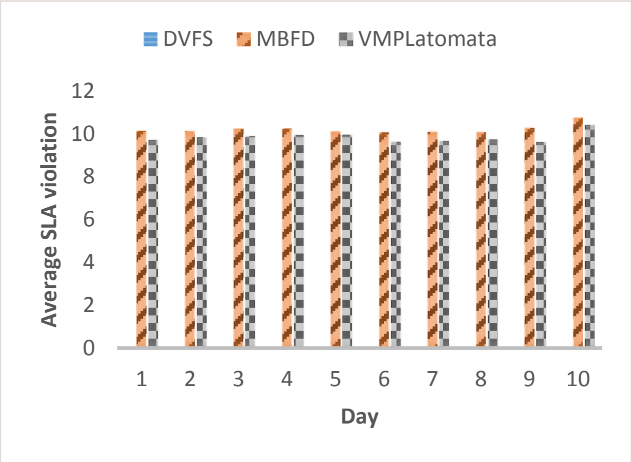

Figure 5. Average SLA violation in the three approaches

Conclusion

In recent years, increasing use of cloud services has caused growth in the desire to develop more data centers. One of the challenging issues in cloud computing environments is high energy consumption in data centers, which has been ignored in the corporate competition to develop data centers. High energy costs and greenhouse gas emissions are among the problems that have emerged as a result of using large data centers, so researchers are now struggling to find an effective approach to decrease their energy consumption.

In this paper, we presented an approach that uses learning automata. We used two parameters, included minimum energy and minimum correlation coefficient, to reduce energy consumption and avoid SLA violation.

References

[1] Brown, R “Report to congress on server and data center energy efficiency: Public law 109-431,” *Lawrence Berkeley National Laboratory*, 2008 .

[2] Gandhi ,A, , H.Balter, R. Das, C.Lefurgy, “Optimal Power Allocation in Server Farms,” *Proceedings of the eleventh international joint conference on Measurement and modeling of computer systems*, pp.157-168, ACM New York, 2009. Article (CrossRef Link)

[3] Peer1 hosting site puts a survey on “Visualized: ring around the world of data center power usage”. From *www.engadget.com.*

[4] Kaplan ,J, Forrest W, Kindler, N, “Revolutionizing Data Center Energy Efficiency,” *McKinsey & Company*, Tech. Rep, 2010.

[5] Johnson P, Marker ,T, “Data centre energy efficiency product profile,” *Tech. Rep*, 2009/05, 2009.

[6] Liu ,L., Wang H., Liu, X. Jin, W. He, Q. Wang, and Y. Chen, “GreenCloud: a new architecture for green data center,” in *Proceedings of the 6th international conference industry session on Autonomic computing and communications industry session*, pp. 29–38,2009. Article (CrossRef Link)

[7] D. Brooks , M. Martonosi, “Dynamic thermal management for high-performance microprocessors,” in The Seventh International Symposium on High-Performance Computer Architecture, pp,. 171–182, 2001. Article (CrossRef Link)

[8] Ye K., Huang D., Jiang X., Chen H., Wu S., Virtual Machine Based Energy-Efficient Data Center Architecture for Cloud Computing: A Performance Perspective, IEEE/ACM International Conference on Green Computing and Communications, on Cyber, Physical and Social Computing, 2010. Article (CrossRef Link)

[9] Kansal A., Zhao F., Liu J., Kothari N., Bhattacharya A.,Virtual Machine Power Metering and Provisioning, Proceedings of the 1st ACM symposium on Cloud computing Pages 39-50, 2010. Article (CrossRef Link)

[10] Beloglazov A., Buyya R.,, “Energy Efficient Resource Management in Virtualized Cloud Data Centers” 10th IEEE/ACM International Conference, 2010. Article (CrossRef Link)

[11] Barroso L. A., Hölzle U., "The datacenter as a computer: An introduction to the design of warehouse-scale machines," *Synthesis Lectures on Computer Architecture*, vol. 4, pp. 1-108, 2009. Article (CrossRef Link)

[12] Barham P., Dragovic B., Fraser K., Hand S., Harris T., Ho A., R.Neugebauery, I. Pratt, and A. Warfield, “Xen and the art of virtualization,” in Proceedings of 19th ACM Symposium on Operating Systems Principles, SOSP’03, 2003. Article (CrossRef Link)

[13] Sotomayor B., Montero R., M. Llorente I., Foster I., , "Virtual Infrastructure Management in Private and Hybrid Clouds",*IEEE Internet Computing*, vol.13, no. 5, pp. 14-22, September/October 2009, Article (CrossRef Link)

[14] Clark C., Fraser K., Hand S., Hansen J.G., Jul E., Limpach C., Pratt I., Warfield A., “Live migration of virtual machines,” In *Proceedings of the 2nd Symposium on Networked Systems Design and Implementation*, NSDI, USENIX, Boston, MA, USA, 2005.

[15] Pathan N., Shetty B.,” Virtual Machine Placement in Cloud”, *International Journal of Computer Science and Information Technologies*, (IJCSIT), Vol. 5 (2) , pp. 1833-1835, 2014

[16] Gupta R., Pateriya R. K., “Survey on Virtual Machine Placement Techniques in Cloud Computing Environment,” *International Journal on Cloud Computing: Services and Architecture (IJCCSA)*, vol. 4, no. 4, 2014.

[17] Bonde D., “Techniques for Virtual Machine Placement in Clouds”, *Master of Technology Report*, 2011.

[18] A. Beloglazov and R. Buyya, “Optimal online deterministic algorithms and adaptive heuristics for energy and performance efficient dynamic consolidation of virtual machines in cloud data centers,” *Concurrency and Computation: Practice and Experience (CCPE)*, 2012, Article (CrossRef Link)

[19] Hieu Trong V., Soonwook H., “A Traffic And Power-Aware Algorithm For Virtual Machine Placement In Cloud Data Center”, *International Journal Of Grid & Distributed Computing*, vol. 7 no. 1, pp. 21-32. 12p., 2014. Article (CrossRef Link)

[20] Panigrahy R., Talwar K., Uyeda L., and Wieder U., “Heuristics for Vector BinPacking”, Microsoft Research Report, 2011.

[21] Xu J.,José A, Fortes B., “Multi-objective Virtual Machine Placement in Virtualized Data Center Environments,” *IEEE/ACM International Conference on Green Computing and Communications*, 2010. Article (CrossRef Link)

[22] Xiaoli W., Zhanghui L., “An energy-aware VMs placement algorithm in Cloud Computing environment,” *Second International Conference on Intelligent System*, 2012.

[23] Goudarzi H., Pedram M., “Energy-Efficient Virtual Machine Replication And Placement In A Cloud Computing System,” *IEEE Fifth International Conference On Cloud Computing*, 2012. Article (CrossRef Link)

[24] Kord N., Haghighi H., “An Energy-Efficient Approach for Virtual Machine Placement in Cloud Based Data Centers,” *5th Conference on Information and Knowledge Technology (IKT)*, 2013. Article (CrossRef Link)

[25] Singh A., Hemalatha M., “Cluster Based Bee Algorithm For Virtual Placement In Cloud Data Centre,” *Journal Of Theoretical And Applied Information Technology*, 30th Nov. 2013.

[26] Hasanul Ferdaus M., Murshed M., Rodrigo Calheiros N, Buyya R., “Virtual Machine Consolidation in Cloud Data Centers using ACO Metaheuristic,” *Lecture Notes in Computer Science*, vol. 8632, pp. 306-317, 2014.

[27] Feller E., Rilling L., Morin Ch., “Energy-Aware Ant Colony Based Workload Placement in Clouds,” *Grid Computing (GRID)*, 12th IEEE/ACM International Conference on , 2011.

[28] Jiang D., Huang P., Lin P., Jiang J., “Energy Efficient VM Placement Heuristic Algorithms Comparison For Cloud With Multidimensional Resources”, Springer-Verlag Berlin Heidelberg, ICICA LNCS 7473, Pp. 413–420, 2012. Article (CrossRef Link)

[29] Rasouli N., Meybodi M. R. Morshedlou,,H,, Virtual Machine Placement in Cloud systems using Learning Automata,, 13th Iranian Conference on Fuzzy Systems (IFSC), 2013. Article (CrossRef Link)

[30] Thathachar M. A. L., Sastry P. S., "Varieties of learning automata: An overview", *IEEE Transactions on Systems, Man and Cybernetics – Part B: Cybernetics*, 32, 6, 2002. Article (CrossRef Link)

[31] N. Moghadasi, M. G. Arani, M. Shamsi. "A Novel Approach for Reduce Energy Consumption in Mobile Cloud Computing," *International Journal of Computer Network and Information Security (IJCNIS)* 7, no. 10, 58, 2015. Article (CrossRef Link)

[32] K. Mogouie, M. G. Arani, M. Shamsi, "A Novel Approach for Optimization Auto-Scaling in Cloud Computing Environment," *International Journal of Modern Education & Computer Science* 7, no. 8 (2015). Article (CrossRef Link)

[33] M. Fallah, M. G. Arani, M. Maeen, "NASLA: Novel Auto Scaling Approach based on Learning Automata for Web Application in Cloud Computing Environment," *International Journal of Computer Application* 117, no. 2, pp. 18-23, 2015. Article (CrossRef Link)

[34] M. Fallah, M. G. Arani. "ASTAW: Auto-Scaling Threshold-based Approach for Web Application in Cloud Computing Environment," *International Journal of U-& E-Service, Science & Technology* 8, no. 3, 2015.

Hossein Ghiasi received the B.S.C degree in Software Engineering from University Payam Noor Aran&Bidgol, Iran in 2008, and M.S.C degree from Azad University of Mahallat, Iran in 2015, respectively. Her research interests include Cloud Computing, Distributed Systems, Cloud Storage Technology and Software Development

Mostafa Ghobaei Arani received the B.S.c degree in Software Engineering from IAU Kashan, Iran in 2009, and M.S.c degree from Azad University of Tehran, Iran in 2011, respectively. He is Currently a PhD Candidate in Islamic Azad University, Science and Research Branch, Tehran, Iran. His research interests include Grid Computing, Cloud Computing, Pervasive Computing, Distributed Systems and Software Development.