International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 4 (2015) pp. 10127-10136

http://www.ripublication.com

A Vikor Method For Distributing Load Balanced Virtual Machine In Cloud Data Center

N.R.Rajalakshmi
Assistant professor
Dept of Computer Science and Engineering
Sacs MAVMM Engineering College,Madurai
rajirajasekaran@gmail.com, 9884712780
Dr.N.Balaji
Professor & Head of Department
Dept. of Information Technology
K.L.N College of Engineering
Sivaganga - 630611

Abstract

In cloud computing, lengthy response time and wasting resources as a consequence of over utilization & under utilization should be avoided to produce better quality of cloud services. To attain sound performance in cloud computing, it is essential to build and sustain a balanced load among the machines. Such as, the load balancing helps in minimizing resource consumption, implementation of failover, enhancing scalability, avoiding bottlenecks and over-provisioning of resources. In this paper, we propose a new model for distributed load balancing allocation of virtual machines in the cloud data center using VIKOR method. Multi criteria decision making model such as VIKOR method is applied to attain healthier load balance in a large scale cloud computing environment.

Keywords: Cloud computing, Load Balancing, VM migration, VIKOR.

Introduction

Ian Foster; Yong Zhao; Ioan Raicu; and Shiyong Lu. (2008) propose that, cloud computing is large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet. Clouds need to run multiple (or even up to thousands or millions of) user applications, and all the applications appear to the users as if they were running simultaneously and could use all the available resources in the cloud.

Susanta Nanda & Tzi-cker Chiueh (2005) state that, virtualization provides an effective solution to the management of data center resources. It is a technology that combines or divides computing resources to present one or many operating environments using methodologies like hardware and software partitioning or aggregation, partial or complete machine simulation, emulation, time-sharing, and many others. Leelipushpam & Sharmila (2013) discuss that, many virtual machines are hosted on the same physical server for optimal resource utilization thereby, reducing the cost of deploying data center. It also enhances the security of physical servers in the data center.

For achieving energy efficiency, load balancing and high availability of physical server in cloud data center, virtual machines are migrated from one physical server to another. Recent research has focused on the globally optimal solution in virtualized data centers. Starting with the work of Almeida, J.; Almeida, V.; Ardagna, D.; Francalanci, C.; Trubian, M. (2006), other research outlined in Khanna, G. ; Beaty, K. ; Kar, G. ; and Koehut, A. (2006), Bobroff, N.; Kochut, A. ; Beaty, K. (2007), Wood, T.; Shenoy, P.; Venkataramani, A.; and Yousif, M. (2007), Wang, X.; LAN, D.; Wang, G.; Fang, X.; Ye, M.; Chen, Y.; and Wang, Q. (2007), Kochut (2008), Hermenier, F.; Lorea, X.; Menaud, J.M.; Muller, G.; and Lawall, J. (2009), and Van and Tran (2009) have also adopted a centralized configuration. Nevertheless, the global optimal solution would create a bunch of unnecessary migrations, resulting in dissipation of the resource.

Here, we propose a novel strategy to give out load balanced physical machine in the cloud data center using VIKOR method which is the multi criteria decision making (MCDM) technique. This method can find the most suitable physical machine PM in the data center for migrating virtual machine (VM). Outcomes show that, our system can achieve better load balancing in large scale cloud computing environment.

System Architecture

The system designed as a large-scale data center consisting of many heterogeneous physical nodes. Each node is characterized by the CPU performance defined in Million Instructions Per Second (MIPS), amount of RAM, network bandwidth and, disk storage. Each node in the data center runs a module of the VM monitor which observes the local resource usages of the node. If the local observations reveal an anomaly that the resources are over-utilized or under-utilized, there are two decisions which need to make, namely:

1. Which VM to migrate from the problematic PM.
2. Which PM has been chosen for migrating VM.

In the next subsection, we will discuss how to choose a VM to migrate from the list of VMs that schedule on the problematic PM and how to choose a PM from the data center to migrate the selected VM.

Virtual Machine Selection

After identifying anomaly, the system chooses the VM from the problematic PM. That is where one or more VMs are chosen from the list of VMs that run on the problematic PM to be migrated to another PM to resolve the anomaly. If the resources are over-utilized, there are one or more VMs need to be migrated. If the resources are under-utilized, all VMs on the problematic PM will be migrated to other PMs in the data center. If all VMs are migrated, then the PM can be switched off.

If the resources are still over-utilized after the migration of a VM, then system sorts all VMs on the problematic PM in decreasing order of current utilization first, and then the system chooses the VM which has the highest utilization in the decreasing order. If the resources are still over-utilized after the migration of the highest utilized VM, then the system chooses the next highest utilization of VM in the decreasing order until the anomaly is resolved.

PM Selection

When the choice of VM is finished, the system starts to select the PM from the data center to transfer the chosen VM to. If there is no PM available in the data center to host the VM, then no migration is happening. Otherwise, all of the PMs in the data center that can host the VM without exceeding the resource threshold compose a set, and the system will choose the most suitable PM from the set using the VIKOR method.

Vikor Method

The VIKOR method began in the form of Lp-metric, which was used as an aggregating function in a compromise programming method and developed into the multi criterion measure for compromise ranking.

The broker uses this method to rank each resource in the data center. We assume a MCDM problem with m nodes or alternatives $A_1, A_2, \ldots A_i \ldots\ldots A_m$, n decision criteria/attributes, $C_1, C_2, \ldots\ldots C_j \ldots\ldots, C_n$ .Each alternative or node is evaluated with respect to n criteria /attributes. w_j is the weight of the jth criterion, expressing the relative importance of the criteria. The rating (performance score) of the jth criterion is denoted by f_{ij} for alternative A_i. The form of Lp-metric is formulated as follows:

$$L_i^p = \{\sum_{j=1}^{n}[w_j(|f_j^* - f_{ij}|)/(|f_j^* - f_j^-|)]^p\}^{1/p}, \qquad 1 \le p \le \infty;\ i = 1, 2, \ldots, m. \qquad (1)$$

The VIKOR method is not only generated with the above form of L_p-metric but it also uses

$L_i^{p=1}$ (as S_i in eq. (2))and $L_i^{p=\infty}$ (as Q_i in eq. (3)) to formulate the ranking measure.

$$S_i = L_i^{p=1} = \sum_{j=1}^{n}[w_j(|f_j^* - f_{ij}|)/(|f_j^* - f_j^-|)] \qquad (2)$$

$$Q_i = L_i^{p=\infty} = \max_j \{ w_j (| f_j^* - f_{ij} |)/(| f_j^* - f_j^- |) \mid j = 1, 2, ..., n \} \quad (3)$$

When p is small, the group utility is emphasized (such as $p=1$) and as p increases, the individual regrets/gaps receive more weight. In addition, the compromise solution $\min_i L_i^p$ will be chosen, its value is closest to the ideal/aspired level. Therefore, in $\min_i S_i$ and $\min_i Q_i$, $\min_i S_i$ express to minimize the sum of the individual regrets/gaps and $\min_i Q_i$ express to minimize the maximum individual regret. In other words, $\min_i S_i$ emphasizes the maximum group utility, whereas $\min_i Q_i$ emphasizes selecting minimum among the maximum individual regrets. Grounded along the above concepts, the compromise ranking algorithm VIKOR consists of the next steps.

1. Determine the best f_j^*, and the worst f_j^- values of all criterion functions $j = 1, 2, ..., n$. If we assume the j^{th} function represents a benefit, then $f_j^* = \max_i f_{ij}$ (or setting an aspired level) and $f_j^- = \min_i f_{ij}$ (or setting a tolerable level). Alternatively, if we assume the j^{th} function represents a cost/risk, then $f_j^* = \min_i f_{ij}$ (or setting an aspired level) and $f_j^- = \max_i f_{ij}$ (or setting a tolerable level).

$$\textit{Alternatives}\quad \begin{array}{c} \\ \\ A_1 \\ \vdots \\ A_i \\ \vdots \\ A_m \\ \\ \\ \end{array} \begin{array}{c} \textit{criteria} \\ \begin{array}{ccccc} c_1 & \cdots & c_j & \cdots & c_n \end{array} \\ \begin{bmatrix} f_{11} & \cdots & f_{1j} & \cdots & f_{1n} \\ \vdots & & \vdots & & \vdots \\ f_{i1} & \cdots & f_{ij} & \cdots & f_{in} \\ \vdots & & \vdots & & \vdots \\ f_{m1} & \cdots & f_{mj} & \cdots & f_{mn} \end{bmatrix} \\ \begin{array}{ccccc} f_1^* & \cdots & f_j^* & \cdots & f_n^* \\ f_1^- & \cdots & f_j^- & \cdots & f_n^- \end{array} \end{array} \quad \begin{array}{c} \textit{normalized} \\ \Rightarrow \\ \times w_j \end{array} \quad \begin{array}{c} A_1 \\ \vdots \\ A_i \\ \vdots \\ A_m \end{array} \begin{array}{c} \textit{criteria} \\ \begin{array}{ccccc} c_1 & \cdots & c_j & \cdots & c_n \end{array} \\ \begin{bmatrix} w_1 r_{11} & \cdots & w_j r_{1j} & \cdots & w_n r_{1n} \\ \vdots & & \vdots & & \vdots \\ w_1 r_{i1} & \cdots & w_j r_{ij} & \cdots & w_n r_{in} \\ \vdots & & \vdots & & \vdots \\ w_1 r_{m1} & \cdots & w_j r_{mj} & \cdots & w_n r_{mn} \end{bmatrix} \end{array}$$

(Original data) (Normalized data)

where, $r_{ij} = (| f_j^* - f_{ij} |)/(| f_j^* - f_j^- |)$, f_j^* is the aspired/desired level and f_j^- is tolerable level for each criterion.

2. Moreover, the weights of criteria, expressing the decision maker´s preference.

2.1) The Coefficient of Variation (CV) is engaged to allocate weights when no preference existed among the criteria considered for resource selection. The Coefficient of Variation (CV) is used to allocate the weights of different criteria. Range standardization was done to transform different scales and units among various criteria into common measurable units.

$$f'_{ij} = \frac{f_{ij} - \min\limits_{1\le j\le n} f_{ij}}{\max\limits_{1\le j\le n} f_{ij} - \min\limits_{1\le j\le n} f_{ij}} \quad (4)$$

$D' = (f')_{m*n}$ is the matrix after range standardization; $max\ f_{ij}$, $min\ f_{ij}$ are the maximum and the minimum values of the criterion(j) respectively, all values in D' are $(0 \le f'_{ij} \le 1)$ .So, according to the normalized matrix $D' = (f')_{m*n}$,the Standard Deviation (σ_j) is calculated for every criterion independently as shown in Equation (5):

$$\sigma_j = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(f'_{ij} - \overline{f'}_j)^2} \quad (5)$$

where $\overline{f'}_j$ is the mean of the values of the j^{th} criterion after normalization and $j = 1, 2\ldots n$ after calculating (σ_j) for all criteria, the (CV) of the criterion (j) will be as shown in equation (6)

$$CV_j = \sigma_j / \overline{f'}_j \quad (6)$$

The weight w_j of the criterion can be defined as

$$w_j = CV_j \Big/ \sum_{j=1}^{n} CV_j \quad \text{where } j = 1, 2\ldots n \quad (7)$$

2.2) The relative importance of the criteria is considered when preference existed among the criteria considered for resource selection. The weights of relative importance of the attributes may be assigned using AHP (Analytic Hierarchy Process) .The steps are enlightened below in the following way:

i. A pair-wise comparison matrix has to be constructed using a scale of relative importance to find out the relative importance of different attributes with respect to the objective. The judgments are entered using the fundamental scale of the AHP. An attribute compared with itself is always assigned with the value 1, so the main diagonal entries of the pair-wise comparison matrix are all 1. The numbers 3, 5, 7, and 9 correspond to the verbal judgments “moderate importance”, “strong importance”, “very strong importance”, and “absolute importance” (with 2, 4, 6, and 8 for compromise between the previous values). Assuming n attributes, the pair-wise comparison of attribute i with attribute j yields a square matrix A_{n*n}

where a_{ij} denotes the comparative importance of attribute *i* with respect to attribute *j*. In the matrix, $a_{ij} = 1$ when $i = j$ & $a_{ji} = 1/a_{ij}$.

ii. We need to know the vector $w = [w_1, w_2, w_3 \ldots w_n]$ which indicates the weight that each criterion is given in pair-wise comparison matrix A.The following method is followed to recover the vector *w* from *A*.
For each of the A's column divide each entry in column i of A by the sum of the entries in column i.This yields a new matrix, called $A - norm$ (for normalized) in which the sum of the entries in each column is 1. Estimate w_i as the average of the entries in row i of $A - norm$.

3. By using the weights of criteria, Compute the values S_i & Q_i, $i = 1, 2, \ldots\ldots\ldots m$ as shown in equation 8 and 9.

$$S_i = \sum_{j=1}^{n} w_j r_{ij}, \tag{8}$$

$$Q_i = \max_j \{w_j r_{ij} \mid j = 1, 2, ..., n\}, \tag{9}$$

4. **C**ompute the index values R_i, $i = 1, 2, \cdots, m$, using the relation

$$R_i = v(S_i - S^*)/(S^- - S^*) + (1 - v)(Q_i - Q^*)/(Q^- - Q^*) \tag{10}$$

Where $S^* = \min_i S_i$ (or setting the best $S^* = 0$), $S^- = \max_i S_i$ (or setting the worst $S^- = 1$), $Q^* = \min_i Q_i$ (or setting the best $Q^* = 0$), $Q^- = \max_i Q_i$ (or setting the worst $Q^- = 1$), and $0 \le v \le 1$ where *v* is introduced as a weight for the strategy of maximum group utility, where as $1 - v$ is the weight of the individual regret. In other words, when $v > 0.5$, this represents a decision making process that could use the strategy of maximum group utility (i.e., if v is big, group utility is emphasized), or by consensus when $v \approx 0.5$, or with veto when $v < 0.5$.

5. Rank the alternatives, sorting by the value of { S_i, Q_i and $R_i \mid i = 1, 2, ..., m$ }, in decreasing order. Propose as a compromise the alternative ($A^{(1)}$) which is ranked first by the measure $\min\{R_i \mid i = 1, 2, ..., m\}$ if the following two conditions are satisfied:

C1.Acceptable advantage: $R(A^{(2)}) - R(A^{(1)}) \ge 1/(m-1)$, where $A^{(2)}$ is the alternative with second position in the ranking list by R; m is the number of alternatives.

C2.Acceptable stability in decision making: Alternative $A^{(1)}$ must also be the best ranked by { S_i or/and $Q_i \mid i = 1, 2, ..., m\}$.If one of the conditions is not satisfied, then a set of compromise solutions is proposed, which consists of:

- Alternatives $A^{(1)}$ and $A^{(2)}$ if only condition C2 is not satisfied.
- Alternatives $A^{(1)}, A^{(2)}, ..., A^{(M)}$ if condition C1 is not satisfied. $A^{(M)}$ is

determined by the relation $R(A^{(M)}) - R(A^{(1)}) < 1/(m-1)$ for maximum M (the positions of these alternatives are close.

The compromise solution is determined by the compromise-ranking method; the obtained compromise solution could be accepted by the decision makers because it provides maximum group utility of the majority (represented by min S, eq. (8)), and minimum individual regret of the opponent (represented by min Q, eq. (9)).

Then, the system migrates VM to the better alternative PM which is chosen by using VIKOR. Likewise, when there are other VMs need to be migrated; the system chooses the most suitable PM for the migrated VMs using the VIKOR method.

Evaluations

Experimental Setup

Here, the simulation framework cloudsim toolkit has been chosen to evaluate the proposed method. We have extended the framework in order to enable load balancing simulation. The simulated data center consists of 200 heterogeneous physical nodes. Each node is modeled to have one CPU core with performance equivalent to 1000, 2000 or 3000 MIPS, 6, 7 or 8 GB of RAM and 2, 3 or 4 TB of storage. The heterogeneous migrating VMs that have one CPU core with performance equivalent to 250, 500, 750 or 1000 MIPS, 3 GB of RAM and 1 TB of storage, and fill the full capacity of the data center. In the system, there are three criteria to evaluate the quality of each alternative in the set. The three criteria are CPU usage, memory usage and bandwidth usage, and each criteria is assigned by a weight. We have also implemented First Fit (FF) and First Fit Decreasing (FFD) methods to compare with proposed VIKOR resource allocations.

Experiment Results

The simulations have been run on each method. Each experiment is repeated for several times and every test result comes from the arithmetic average of experimental values. We consider two thresholds in the system: the lower and upper utilization thresholds. The simulation results are presented in Figure 1-2. Figure1 shows the percentages of SLA violations by the VIKOR method when the lower utilization threshold is from 10% to 40% and the upper utilization threshold is from 60% to 90%. We have defined that, SLA violation occurs when a VM cannot get amount of resources that are needed. The results show that, with the increase of the lower utilization threshold and the decrease of the upper utilization threshold, the percentage of SLA violations increases. The VIKOR method can reduce a lot of the percentage of SLA violations. Thus, the VIKOR method can make the system to provide better QoS.

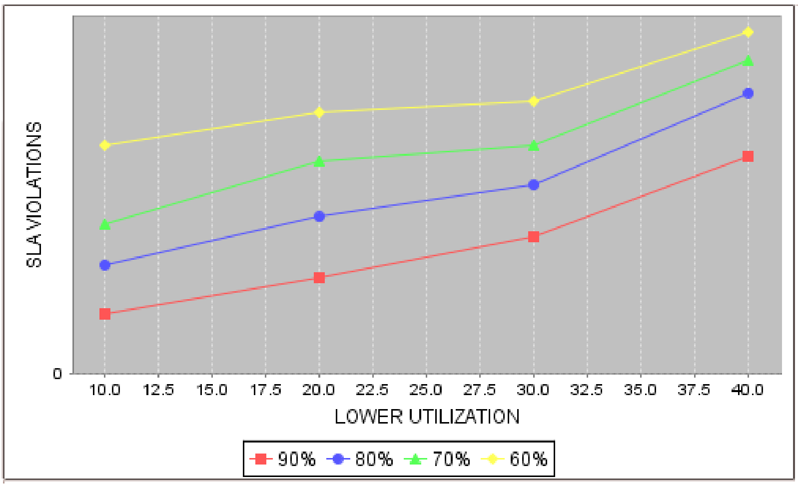

Figure 1: SLA violations by the VIKOR method for different values of the utilization thresholds

We have used 10% as the lower utilization threshold and 90% as the upper utilization threshold. The vikor method has the least percentage of SLA violations comparative with FFD and FF methods. .

Figure 2 shows the number of migrated VM provisioned by the three methods first fit, first fit decreasing and VIKOR method. The VIKOR method hosts the highest number of virtual machines on physical machines in the data center during simulations. Experimental results show that, VIKOR method provisions the resource for more number of migrated virtual machines compared to FFD and FF. The reason is that the FFD and FF method are the globally optimal solution which makes a lot of unnecessary migrations, but the system by the V IKOR method makes less migration. Less migration can make the system more stable.

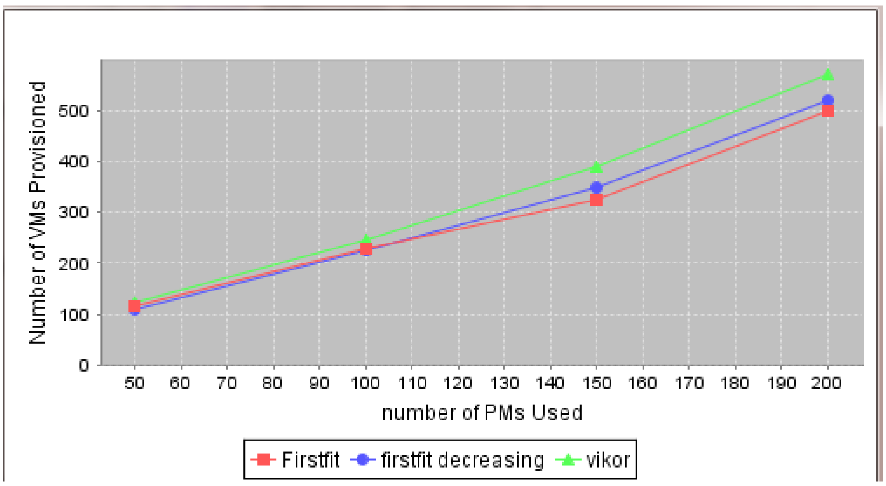

Figure 2: The number of VM migrations and the number of PM used by the three methods

Conclusion

We have proposed a VIKOR model for distributed load balancing allocation of virtual machines in the cloud data center. Results show that our system can attain better load balancing in a large-scale cloud computing environment with less VM migration.

Reference

[1] Hadi Goudarzi and Massoud Pedram(2011),“Multi-dimensional SLA-based Resource Allocation for Multi-tier Cloud Computing Systems”, in Proceedings of IEEE Conference on Cloud Computing(CLOUD) , Washington, pp.324-331.

[2] Sotomayor, B.; Montero, R.S.; Llorente, I.M.; and Foster, I.(2008), “Capacity Leasing in Cloud Systems using the OpenNebula Engine”, Workshop on Cloud Computing and its Applications, Chicago,Illinois, USA.

[3] Srikantaiah, S.; Kansal, A.; and Zhao, F.(2008), “Energy aware consolidation for cloud computing Power Aware Computing and Systems”, San Diego, USA.

[4] Komal Singh Patel ; Sarje, A.K.(2012),“VM Provisioning Method to Improve the Profit and SLA Violation of Cloud Service Providers, in Proceedings of IEEE Conference on Cloud Computing in Emerging Markets(CCEM),Bangalore,India, pp.168-172.

[5] Ian Foster ; Yong Zhao ; Ioan Raicu ; and Shiyong Lu.(2008), “ Cloud Computing and Grid Computing 360-Degree Compared”,Grid Computing environment workshops(GCE’08),Austin,Tx,pp.1-10.

[6] Susanta Nanda and Tzi-cker Chiueh,(2005),“A Survey on Virtualization Technologies”,Experimental Computer System lab.

[7] Appleby, K.C. ; Fakhouri, S. ; Fong, L. ; Goldszmidt, G. ; Krishnakumar, S. ; Pazel, D.P. ; Pershing, J. ; Rochwerger, B. (2001),“Oceano-SLA-based management of a computing utility”, in: Proceedings of the IFIP/IEEE Symposium on Integrated Network Management, pp.855-868.

[8] Leelipushpam, P.G.J. ; Sharmila, J. (2013), “Live Migration Techniques in Cloud Environment-A Survey”,in Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013),pp.408-411.

[9] Almeida, J.; Almeida, V.; Ardagna, D.; Francalanci, C.; Trubian, M.(2006),"Resource Management in the Autonomic Service-Oriented Architecture," in ICAC '06: IEEE International Conference on Autonomic Computing, pp.84-92.

[10] Khanna, G. ; Beaty, K. ; Kar, G. ; and Koehut, A.(2006),"Application Performance Management in Virtualized Server Environments," in Network Operations and Management Symposium, NOMS 2006, 10th IEEE/IFIP, pp.373-381.

[11] Bobroff, N.; Kochut, A. ; Beaty, K.(2007), "Dynamic Placement of Virtual Machines for Managing SLA Violations," in IM '07: 10th IFIP/IEEE

International Symposium on Integrated Network Management, 2007,pp.1l9-128.

[12] Wood, T. ; Shenoy, P. ; Venkataramani, A. ; and Yousif, M.(2007),"Black-Box and Gray-Box Strategies for Virtual Machine Migration," in 4th USENIX Symposium on Networked Systems Design and Implementation, 2007,pp. 229-242.

[13] Wang, X. ; Lan, D. ; Wang, G. ; Fang, X. ; Ye, M. ; Chen, Y. ; and Wang, Q. (2007),"Appliance-Based Autonomic Provisioning Framework for Virtualized Outsourcing Data Center," in Autonomic Computing, ICAC'07-Fourth International Conference , 2007, pp. 29-29.

[14] Kochut, A. (2008), "On Impact of Dynamic Virtual Machine Reallocation on Data Center Efficiency," in Proceedings of IEEE MASCOTS, 2008,pp.13-20.

[15] Hermenier, F. ; Lorea, X. ; Menaud, J.M. ; Muller, G. ; and Lawall, J. (2009),"Entropy: A Consolidation Manager for Clusters", in VEE '09:Proceedings of the 2009 ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments, pp. 41-50.

[16] Van, H.N. and Tran, F.D.(2009), "Autonomic Virtual Resource Management for Service Hosting Platforms," in ICES '09: Proceedings of the International Conference on Software Engineering Workshop on Software Engineering Challenges of Cloud Computing. IEEE Computer Society, 2009, pp. 1-8.

[17] Buyya, R. ; Ranjan, R. ; and Calheiros, R.N.(2009), "Modeling and simulation of scalable cloud computing environments and the CloudSim toolkit: Challenges and opportunities." In Proceedings of the 7th High Performance Computing and Simulation Conference (HPCS'09). IEEE Press, NY, USA.