An Allocation Method for Resource of IaaS Platform with Energy-saving Consideration in the Data Center

Qingxin XIA

School of Computer Science and Engineering
Beihang University, BUAA
Peking, China
xiaqingxin@cse.buaa.edu.cn

Yuqing LAN, Long WANG, Limin XIAO, Xiang WANG

School of Computer Science and Engineering
Beihang University, BUAA
Peking, China
lanyuqing@buaa.edu.cn

Abstract—Reducing energy consumption and maximizing resource efficiency have become two important issues without violating service level agreements(SLAs) on Infrastructure as a Service(IaaS) clouds. For the important issues, using the method of resource classification, this study proposes a resource provisioning method considering on energy-efficient for allocating VMs to physical machines(PMs) for the first time. Under the premise of ensuring performance and energy-saving target, through the analysis of the main factors to affect the energy consumption and the migration process of VM, the first VM resource allocation ultimately makes the use of resource to maximize and makes less hosts allocated. In the evaluation process, based on CloudSim 3.03 toolkit, it is extended to construct the implement of this research. What's more, using real-workload(PlanetLab) carried out experiments simulating. Experimental results show that the proposed approach outperforms other proposed ones in energy consumption.

Keywords-Energy-saving; IaaS; Resource allocation; ANN; Classification

I. INTRODUCTION

Cloud computing providers provide services and infrastructures to thousands of users, enabling them to have full control in the provided IaaS clouds. What's more, allocation, re-allocation and migration of VMs must be a series of programs which are designed in the practical applications. However, even in the state-of-threat data centers, the massive amount of PMs leads to high energy consumption and carbon footprint of the data center, as well as high operation costs[1]. Moreover, it is not surprising to become a hot research in this area.

There are several different possibilities to save energy constructing cloud computing infrastructures in the early period. On the one hand, through network partitioning perspective, solving deployment problem of computing resource makes switches and PMs have a unified division for the energy-saving in the data center[2]. But the relevant research has not combined with some IaaS platform strategy study so that it can not obtain ideal results. On the other hand, the allocation of VMs to PMs can waste a lot of energy consumption, if the resources allocated to these VMs are overestimated. This is usually the case, as SLAs, what the size of a VM is normally based upon, are often specified rather pessimistically to guarantee for peak workload. Furthermore, powering off the exact number of unused PMs is a major approach to enhance the level of energy-efficient. Consequently, finding usable solutions on the IaaS platform, optimizing the first-allocation of data center's resources, is a series of important problems to be solved.

The research presented in this paper unifies all the three mentioned methodologies (data center division, VM first-allocation, PM power management), presents concrete implementations for each of them, and tackles all of the mentioned constraints (PM division, VM workload type, and power management costs). The main contributions of this paper are:

1. By analyzing two major factors of impacting energy consumption, proposing the energy-aware approach of first-allocation on the IaaS clouds according to VM resources classification using SLA information and workload.

2. Solving an issue that VM resources to consume higher energy consumption can not been deployed on the PMs with the higher performance-energy ratio, making data center waste additional energy consumption.

II. RELATED WORK

We can divide related work in this area into two parts: resource-efficient VM management, and energy-efficient PM management. As to the first part, Petrucci[4] and Bichler[5] investigate one general resource constraint, whereas we focus on analyzing the causes of impacting energy consumption on the IaaS. The Sandpiper framework[6] offers black-box and gray-box resource management for VMs. Contrary to his approach, the first-allocation approach this study proposed emphasizes resource allocation strategy. Also Yuan J et al. [7] with their model pursue SLA violation minimization, but can only execute one action per iteration and neglect the energy consumption of executed actions. Additionally, Maurya K et al. [8], none of the presented papers uses a CPU-and-memory customized and SLA-based approach for possible reconfiguration and self-adaptation. Hoyer et al. [9] also undertake a speculative approach but by overbooking PM resources. They assign VMs to PMs that would exceed their maximum resource capacities, because VMs hardly ever use all their assigned resources. In order to realize this allocation they also take workload correlation of

different VMs into consideration . Our approach is to analyze resources as classification results to allocate to most suitable PMs. Stillwell et al. [10] in a similar setting define the resource allocation problem for static workloads, present the optimal solution for small instances and evaluate heuristics by simulations. Nathani et al. [11] also deal with VM placement on PMs using scheduling techniques. [3,12] reacts to changing workload demands by starting new VM instances, taking into account VM startup time, they use prediction models to have VMs available already before the peak occurs. In a word, we can say that there has been a great deal of work on the different action levels, whereas VM first-allocation has not been observed yet.

As to energy-efficient PM management, especially rule-based systems have gained some interest. Paschke [13] et al. investigate a rule-based approach in combination with the logical formalism ContractLog. Their formalism specifies rules to trigger after a violation has occurred, but it does not deal with avoidance of SLA violations. As far as the second level is concerned, several papers focus on different levels. [14] focus on VM migration and on turning on and off physical machines, whereas our paper achieves a more holistic approach taking all these mentioned levels and VM reconfiguration into account. Feller et al. [15, 16] tackle the cost of live migration of VMs regarding the response time of the services inside the VMs in order to match the response time with the SLA requirements of the services. Liu et al. [17] also have studied live migration of virtual machines in order to model the performance and energy of the migration. The architectural framework proposed in [18] for green clouds also achieves VM reconfiguration, allocation and reallocation. Our research provides a more wholesome approach than related work and integrates most of the different possible action levels seen in the literature.

III. THE SYSTEM MODEL AND POWER MODEL

A. The system model

In this paper, the targeted system is an IaaS environment, represented by a large-scale three-tier Ethernet data center consisting of M heterogeneous physical nodes(PMs). Each node i is characterized by the CPU performance defined in Millions Instructions Per Second (MIPS), amount of RAM and network bandwidth. The PMs do not have local disks, the storage is provided as a Network Attached Storage (NAS) to enable live migration of VMs.
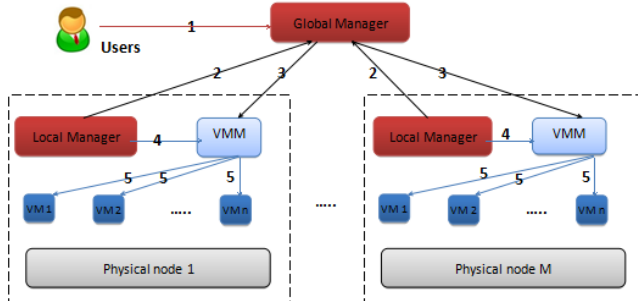

Figure 1.   The system model.

The software layer of the system is tiered comprising local and global managers (seeing Fig. 1). The local managers reside on each node as a module of the VMM. Their objective is the continuous monitoring of the node's CPU utilization, resizing the VMs according to their resource needs, and deciding when and which VMs should be migrated from the node (4). The global manager resides on the master node and collects information from the local managers to maintain the overall view of the utilization of resources (2). The global manager issues commands for the optimization of the VM placement (3). VMMs perform actual resizing and migration of VMs as well as changes in power modes of the nodes (5).

B. The power model

Power consumption by computing nodes in data centers is mostly determined by the CPU, memory, disk storage, power supplies and cooling systems. Recent studies have shown that the power consumption by PMs can be accurately described by a linear relationship between the power consumption and CPU utilization, even when Dynamic Voltage and Frequency Scaling (DVFS) is applied. Therefore, instead of using an analytical model of power consumption by a server, we utilize real data on power consumption provided by the results of the SPEC power benchmark.

We have selected four servers with dual-core/four-core/eight-core CPUs published in February 2011 or earlier: HP ProLiant ML110 G4 (Intel Xeon 3040, 2 cores×1860 MHz, 4GB), and HP ProLiant ML110 G5 (Intel Xeon 3075, (2 cores×2660 MHz, 4GB), and HP ProLiant DL120 G5 (Intel Xeon X3360, 4 cores × 2830 MHz, 8GB), and HP ProLiant DL180 G5 (Intel Xeon L5420, 8 cores×2500 MHz, 8GB). The configuration and power consumption characteristics of the selected servers are shown in TABLE IV. The reason why we have not chosen servers with more cores is that it is important to simulate a large number of servers to evaluate the effect of VM consolidation. Thus, simulating less powerful CPUs is advantageous, as less workload is required to overload a server.

IV. THE ALLOCATION METHODOLOGY

A. Data Center Division

Whether from the compatibility or the cost, the selections of hardware for Ethernet tend to be more unanimous, it is usually purchased in the same brand. So taking the general situation, we does not optimize hardware with energy aware, but divide the physical hosts. Three-tier trees of PMs and switches form the most widely used data center architecture . It consists of the core tier at the root of the tree, the aggregation tier that is responsible for routing, and the access tier that holds the pool of PMs.

According to the experimental design, the structure of data center to be divided will be given in the following: the core tier has 4 core switches, the aggregation tier consists of 8 three-tier switches, the access tier consists of 17 two-tier switches with 48 ports, and 800 PMs deployed. Each of these

two-tier switches controls a group of the same type of PMs except the mixed group, Each of these two-tier switches controls a group of the same type of PMs except the mixed group. Nevertheless, how to deploy on PMs in the data center? Usually, the ratio of performance and power is a critical parameter. Now, the principles of energy aware in this study are as followed:

- Power off PMs in idle status to meet the requirement of energy saving.
- Power off the two-tier switches, when all the PMs controlled power off, to realize the requirement of energy saving in the access tier.
- Power off three-tier switches, when all the two-tier switches controlled powered off, to realize the requirement of energy saving in the aggregation tier.

Now, one example of Data center division will be given(seeing Fig. 7).

B. VMs Allocation

When a data center is constructed in the early period, applications of VMs have to be allocated to the suitable PMs with energy-aware consideration. Moreover, after replanning the data center or adjusting applications of VMs, VMs allocation is likewise carried out.

As possible deployment actions, we divide first resource allocation problem into the three following factors:

1) Resource Classification

Considering the energy consumption to deploy VMs, the workload and resource allocation is a pair of important factors of mutual association. The workload reflects CPU utilization, memory usage and other characteristics in a certain period of time after the user purchasing the VMs. Resource allocation is that what type of PM resource the cloud computing providers will deploy applications of VM customers bought on. Our research is focused on providing a solution through the analysis on the workload, to make any workload's or a kind of workloads' VMs can be deployed on the most suitable PMs with energy aware considered. While strategies provided by CloudSim[19], can only ensure the current VM applications to be deployed have the best deployment scheme, but can not guarantee the other all the VM applications get its own best one.

There are two kinds of ways to conduct resource classification action in this study.

- Carrying out deployment of resource classification according to the VM workload
- Carrying out deployment of resource classification according to the SLAs signed with the users

According to SLA classification Method, there hasn't been an ideal method to extract the characteristics specified in SLA, while the granularity of classification got by artificial way is very coarse, which cannot reflect the advantage of saving energy well. This paper will do further research on VM workload, exploring the methods of resource classification deployment in order to save energy. Now according to the first method , the classification model will be introduced based on BP neural network algorithm.
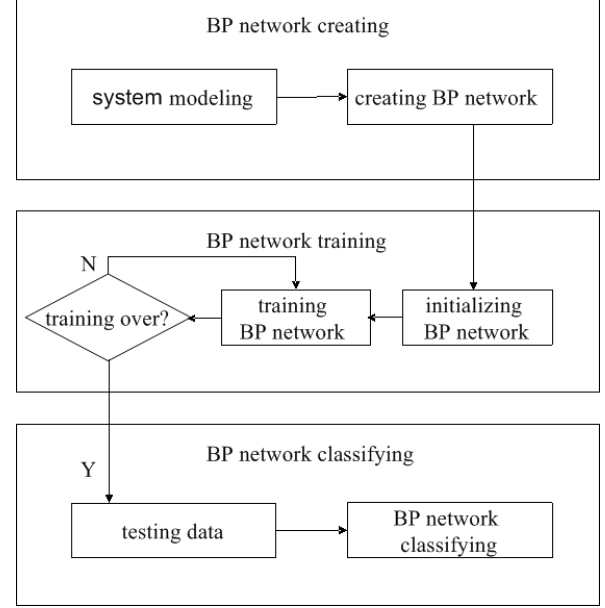

Figure 2.   BP network classifying algorithm flow chart.

The resource classification models based on BP artificial neural network includes three parts: network constructing, network training and network classifying. The algorithm flow chart is shown by Fig. 2.

Feature extracting and pattern recognition of the workload is a set of important method to determine the input and output of BP neural network classification model.

The feature extracting reflects the effective way how the workload changes over a period of time , and the pattern recognition is a classification pattern gotten after matching the features extracted and the reference models expected. The process is(seeing Fig. 3): first, after deploying VM applications, workload data will be obtained over a period of time, then after preconditioning, mathematical method will be used to extract the data of feature, which will be the workload feature model of VM application, which is the input vector X of BP neural network. Then, after comparing the models of workload feature and reference pattern, the best classification model of the workload can be obtained, which is the output vector Y of BP neural network.

According to requirements of SLAs signed with the user, deploying all of the VM applications need three cycles of tests, the workload data of VM application can be obtained, and saved as the corresponding sample, then a set of three samples will be constructed.

The normalization is presented to deal with the workload data in the Preconditioning phase, which is the common method of processing the input data in artificial neural network (ANN).
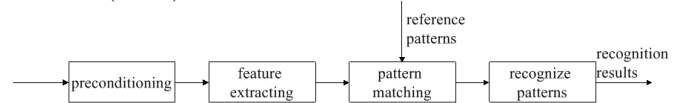

Figure 3.   Feature exacting and pattern recognizing diagram.

The normalization of the workload data change all data into [0, 1], which is in order to reduce the network error caused by magnitude differences among the data in all dimensions. There are mainly two kinds of data normalization methods.

a) Maximum-minimum Method. The function form is as followed:

$$x_k = (x_k - x_{min})/(x_{max} - x_{min}) \qquad (1)$$

In formula (1), x_{min} is the minimum data, x_{max} is the maximum data.

b) Mean Variance Method. The function form is as followed:

$$x_k = (x_k - x_{mean})/x_{var} \qquad (2)$$

In formula (2), x_{mean} is the average data, x_{var} is the variance of data.

In this study, the first data normalization method is taken.

After calculating the mean of workload, lower quartile (Q1) of workload, median of workload and upper quartile (Q3) of workload to be the four features of workload data, the input vector $X = (mean, Q1, median, Q3)$ is combined.Then, pattern matching is happened.

In the training set, each sample must have a desired output, the matching between input and output is ANN schema matching model. For example, $X \sim Y$, and

$$Y = \begin{cases} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{cases}$$

means the classification mode with four kinds of output, so, after recognizing patterns processing, such the result $X' \sim Y' = (mean, Q1, median, Q3) \sim (1,0,0,0)$, which is about input and output matching model can be obtained.

In this way, constructing of BP network can determine the structure of BP network according to the feature extracting and pattern matching of the system's I/O, and determine the dimensions of the input signal according to the characteristics of the workload data, with the numbers of resource workload to be classified as the numbers of output nodes, besides the BP neural network is set to single hidden layer, and further more to determine the structure of BP neural network 4-6-4, namely: the input layer with 4 nodes, the hidden layer with 6 nodes, output layer with 4 nodes.

After creating the classification model, it is needed to train the model. The training data set should have sufficient number (usually more than 1000 groups of training data), with a certain width, enough training frequency, and a certain depth, achieving the effect of convergence, so as to obtain a more accurate classification result. The initialization process of network is mainly to initialize the BP network weights and thresholds, with the random function generating, the initial weights and thresholds can be selected differently, which can only affect the convergence speed, but impact on

the classification results. Finally, classification model training ended, the optimal weights and thresholds will be obtained and stored, the workload data will be classified.

2) PM Management

On the basis of this most basic way of energy saving in the field, combining with the analysis of Resource classification, we know that VM allocation can be deployed on the PMs more effectively when distinguishing workload types, and then the purpose of saving energy consumption can be achieved. So, the factor of analyzing energy aware resource deployment on the IaaS platform is PM management. But how can VM application be distinguished better and deployed on PM in order to save the electricity? This is the purpose of analyzing the factor.

Now, we will construct the trying allocating energy model.

$$E = getPower\left(\frac{avg(U_{workload}) \cdot TotalMips_{VM}}{TotalMips_{PM}} \times 100\%\right) * T \quad (3)$$

In formula (3), E represents the energy consumption during the time T how long VM application deployed on PM, $getPow()$ is linear energy consumption calculation algorithm from Cloudsim, $avg(U_{workload})$ is the mean data of VM workload during the time T, $TotalMips_{VM}$ is total mips of VM type, $TotalMips_{PM}$ is the result of the core numbers multiplied by the frequency in the corresponding PM, T is the time in testing VM workload.

With the trying allocating energy model, the priority in the two factors of resource classification and service time can be judged. It tells which factor will be considered in priority when deploying the application of VM on the divided PMs, making the lowest energy consumption in the global data center. In addition, the trying allocating energy model will go on to be applied on the corresponding VM application deployment in the workload set of resource classification. The detail application can be seen in the algorithm, while the effect of saving energy consumption can be seen in the part of experimental evaluation.

```
Algorithm: Power Aware Resource Classification Allocate
(PARCA)
Input: hostList, vmList   Output: allocation of Vms
foreach vm in vmList do
    foreach host in hostList do
        If host has enough resources for vm then
            vm.power ←
        getpower(vmEven*vmTotalMips/hostTotalMips)*T
//vmList.sortDecreasingVmPower()
vmList.workloadClassifyingVmPower()
foreach vm in vmList do
    minPower←MAX
    allocatedHost←NULL
    foreach host in hostList do
        If host has enough resources for vm then
            power ← getpower(host, vm)
            if power < minPower then
                allocatedHost ← host
                minPower ← power
    If allocatedHost ≠ NULL then
        allocation.add(vm, allocatedHost)
```

V. EVALUATION

A. Experiment Setup

We have simulated the data center that comprises of 800 heterogeneous physical nodes using CloudSim 3.03 toolkit, one quarter of which are HP ProLiant ML110 G4 servers, and one quarter of which are HP ProLiant ML110 G5 servers, and one quarter of which are HP ProLiant DL120 G5 servers, and the other quarter consists of HP ProLiant DL180 G5 servers. The characteristics of the servers and data on their power consumption are given in TABLE IV.

The characteristics of the VM types correspond to Amazon EC2 instance types with the only exception that all the VMs are single-core, which is explained by the fact that the workload data used for the simulations come from single-core VMs. For the same reason the amount of RAM is divided by the number of cores for each VM type: High-CPU Medium Instance (2500 MIPS, 0.85 GB); Extra Large Instance (2000 MIPS, 3.75 GB); Small Instance (1000 MIPS, 1.7 GB); and Micro Instance (500 MIPS, 613 MB).

To make a simulation-based evaluation applicable, it is important to conduct experiments using workload traces from a real system. We have used the PlanetLab data on the CPU utilization by more than one thousand VMs from heterogeneous servers. The interval of utilization measurements is 5 minutes. The data confirms the statement made in the beginning: the average CPU utilization is far below 50%. During the simulations, each VM is randomly assigned a workload trace from one of the VMs. The specific process of the experiment is shown in Fig. 7.

B. Evaluation Metrics

In order to compare the energy costs of the allocation method and other strategies we use several metrics to evaluate their superiority of energy aware. Metric E is the total energy consumption by PMs of the data center caused by the application workloads. Metric m is the number of VM migrations initiated by the VM manager during the adaptation of the VM placement. Metric n is the the number of PMs after VM first allocation, which reflects the deployment efficiency of resource. It means if resource is the same, the lower value means the less PMs to be used. The metric of SLA violation (SLAV) is a combination of two parameters, the details for measuring the level of SLA violations in an IaaS environment are as followed: (1) the percentage of time, during which active hosts have experienced the CPU utilization of 100%, SLA violation Time per Active Host (SLATAH), and (2) the overall performance degradation by VMs due to migrations, Performance Degradation due to Migrations (PDM). The reasoning behind the SLATAH is the observation that if a host serving applications is experiencing 100% utilization, the performance of the applications is bounded by the host capacity. Therefore, VMs are not being provided with the required performance level. We denote the combined metric SLAV, which is calculated as shown as TABLE I.

C. Simulation Results and Analysis

The simulation of this study adopted two scenarios:

Metrics	Description	Explanation
E	the total energy consumption	by PMs of a data center caused by the application workloads
SLAV	the metric of SLA violation	measuring the level of SLA violations : SLAV = SLATAH ·PDM
m	the number of VM migrations	initiated by the VM manager during the adaptation of the VM placement
n	the number of PM after VM allocation	the number of PM after carrying out VM first allocation

scenario A and scenario B. The first scenario applied DVFS energy-saving strategy without migrations, the second scenario applied MAD-MMT energy-saving strategy with migrations. According to the analysis of the experimental results, we might judge whether the deployment method proposed in this paper is more effective in energy saving.

The Median Absolute Deviation (MAD) is the deployment strategy of VMs, measuring the statistical dispersion. The whole algorithm is stated in the CloudSim, and it is compared to other algorithms in the evaluation process. Using the workload data of ten days, the experiment adopted the energy-aware strategy of DVFS that CloudSim provided without migrations. We have simulated all four first allocation algorithms (Random, Bin-packing, Workload classification, Resource classification(PARCA)) to verify the energy-saving advantage by the evaluation metrics(E and n). The results produced by the selected algorithms in the scenario A are shown in Fig. 4.

In Fig. 4, the results show that the proposed allocation algorithm of resource classification(PARCA) has more obvious advantages than other algorithms in the CloudSim by the E(Energy consumption) metric. The same as PARCA uses less PMs to deploy VMs than other algorithms in the CloudSim by the n(the number of PMs after allocation) metric. This is further verified, the allocation method that this study proposed is more efficient in resource allocation.

We can obtain TABLE II after analyzing the evaluation metrics above, the value of which is the mean one of 10 days' workload data.

After analyzing the simulation data in TABLE II, we can conclude that workload classification and PARCA significantly outperform other allocation policies, such as Random and Bin-packing. They can respectively save 11.98% and 42.43% of the electrical energy compared with Bin-packing. About the number of PMs in the first deployment, they increase the rate of optimizing the deployment of resources utilization with 29.49% and 42.40%.

Scenario B adopted the energy-aware strategy of MAD-MMT-2.5 that CloudSim provided. We have simulated there first allocation algorithms (MAD, Workload classification, Resource classification) to verify the energy-saving advantage by the evaluation metrics(E, m and SLAV).

In Fig. 5, the results show that the proposed PARCA has more obvious advantages than other algorithms in the CloudSim by the E metric. The same as PARCA arouses less VM migrations than other algorithms in the CloudSim by the m metric.
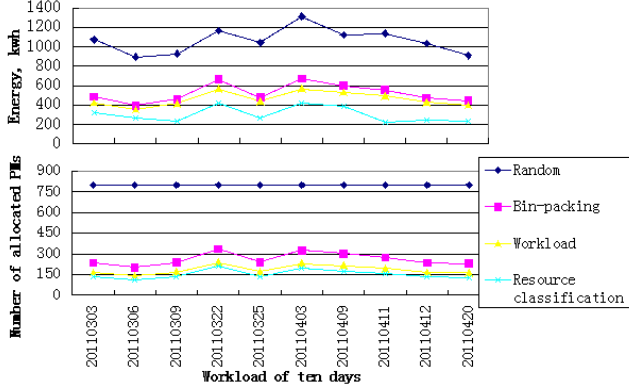
Figure 4. The comparison of four allocation algorithms by E and n metrics in scenario A.
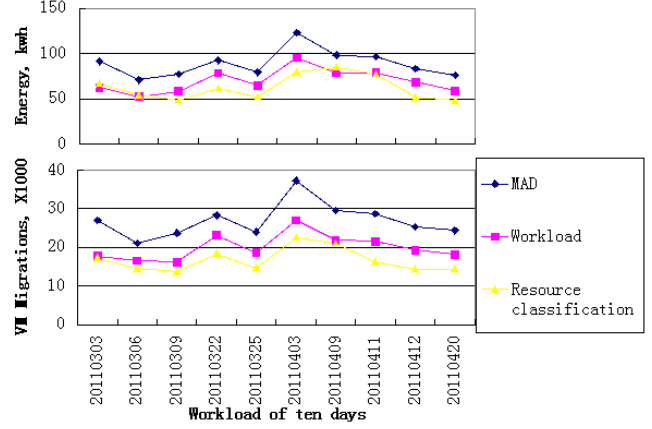

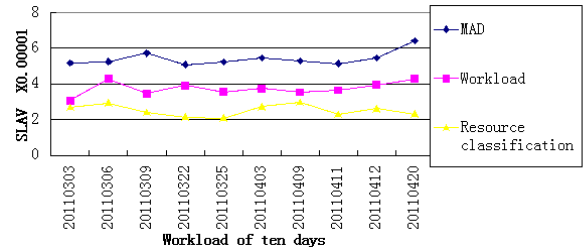Figure 5. The comparison of three allocation algorithms by E and m metrics in scenario B.

Figure 6. The comparison of three allocation algorithms by the SLAV metric in scenario B.

TABLE II.   SIMULATION RESULTS OF ALL THE ALGORITHMS IN THE SCENARIO A(MEAN VALUE)

Policy	Energy(kWh)	n
Random	1061.018	800
Bin-packing	523.08	261
Workload classification	460.423	184
Resource classification	301.152	150

In Fig. 6, the results show that the proposed allocation algorithm of resource classification has the minimum value of SLA violation which is from 2×10^{-5} to 3×10^{-5} than other algorithms in the CloudSim by the SLAV(SLA violation) metric. So the algorithm has enhanced a large amplitude. What's more, the allocation method proposed also has the higher efficiency in resource allocation with migrations.

We can obtain TABLE III after analyzing the evaluation metrics above, the value of which is the mean one of 10 days'workload data. This will help to compare the optimized scope between different algorithms.

After analyzing the simulation data in TABLE III, we can conclude that workload classification and resource classification significantly outperform other allocation policies, such as MAD. They can respectively save 21.65% and 29.71% of the electrical energy compared with MAD. About the number of migrations, 25.65% and 38.01% of them have been decreased. As for the SLAV metric, 30.97% and 53.63% of them have been reduced.

In addition, in order that readers can reproduce the experiment, relative programs and evaluation results of this experiment can be accessed and downloaded on http://wiki.l-cloud.org/ by you.

VI. CONCLUSION

A contribution of this paper is to solve the problem of giving the the highest priority set of resource the physical machine resource with the optimal energy aware, making resources and energy consumption ratio of resource deployment higher and provide better tolerance to IaaS platform. In this paper we have presented a method of the resource allocation with energy-saving consideration for governing Cloud Computing infrastructures to achieve two

TABLE III.   SIMULATION RESULTS OF ALL THE ALGORITHMS IN THE SCENARIO B(MEAN VALUE)

Policy	Energy(kWh)	m(X10^3)	SLAV(X10^{-5})
MAD	88.845	26.945	5.428
Workload classification	69.613	20.034	3.747
PARCA	62.452	16.703	2.517

goals: reducing energy consumption and minimizing migration times. From the experiment, the results have shown the proposed method not only is more efficient. For future work we plan to focus more on a possible heterogeneity of the systems, refining the migration model, and integrating the framework into a real-world Cloud computing environment.

[1]   Peter Johnson, Tony Marker, "Data center energy efficiency product profile", Technical report, 2009.

[2] Dzmitry Kliazovich, Pascal Bouvry, Samee Ullah Khan, "DENS: data center energy-efficient network-aware scheduling", Cluster Comput, DOI 10.1007/s10586-011-0177-4, September 2011.

[3] Quang-Hung N, Nien P D, Nam N H, Tuong N H, Thoai, N. "A genetic algorithm for power-aware virtual machine allocation in private cloud", Springer Berlin Heidelberg, 2013: 183-191.

[4] Vinicius Petrucci, Orlando Loques, and Daniel Mossé "A dynamic optimization model for power and performance management of virtualized clusters", ACM, In e-Energy '10, pages 225‑233, New York, NY, USA, 2010.

[5] Martin Bichler, Thomas Setzer, and Benjamin Speitkamp, "Capacity Planning for Virtualized Servers". Presented at Workshop on Information Technologies and Systems (WITS), Milwaukee, Wisconsin, USA, 2006.

[6] Timothy Wood, Prashant Shenoy, Arun Venkataramani, and Mazin Yousif, "Sandpiper: Black-box and gray-box resource management for virtual machines". Computer Networks, 53(17):2923‑2938, 2009.

[7] Yuan J, Miao X, Li L, Jiang X. An Online Energy Saving Resource Optimization Methodology for Data Center. Journal of Software, 8(8): 1875-1880, 2013.

[8] Maurya K, Sinha R. "Energy Conscious Dynamic Provisioning of Virtual Machines using Adaptive Migration Thresholds in Cloud Data Center", 2013.

[9] Marko Hoyer, Kiril Schr?der, and Wolfgang Nebel, "Statistical static capacity management in virtualized data centers supporting fine grained qos specification", ACM, In Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking, e-Energy '10, pages 51‑60, New York, NY, USA, 2010.

[10] Mark Stillwell, David Schanzenbach, Frederic Vivien, and Henri Casanova, "Resource allocation algorithms for virtualized service hosting platforms", Journal of Parallel and Distributed Computing, 70(9):962‑974, 2010.

[11] Amit Nathani, Sanjay Chaudhary, and Gaurav Somani, "Policy based resource allocation in iaas cloud", Future Generation Computer Systems, 28(1):94‑103, 2012.

[12] Sadeka Islam, Jacky Keung, Kevin Lee, and Anna Liu, "Empirical prediction models for adaptive resource provisioning in the cloud", Future Generation Computer Systems, 28(1):155‑162, 2012.

[13] Adrian Paschke, Martin Bichler, "Knowledge representation concepts for automated SLA management", Decision Support Systems, 46(1):187‑205, 2008.

[14] M. Mazzucco, D. Dyachuk, and R. Deters, "Maximizing cloud providers' revenues via energy aware allocation policies", In CLOUD 2010, pages 131‑138, 2010.

[15] Eugen Feller, Louis Rillingy, and Christine Morin, "Snooze: A Scalable and Autonomic Virtual Machine Management Framework for Private Clouds", informatics mathematics, December 2011.

[16] Eugen Feller, Cyril Rohr, David Margery, and Christine Morin, "Energy Management in IaaS Clouds: A Holistic Approach", informatics mathematics, April 2012.

[17] Haikun Liu, Cheng-Zhong Xu, Hai Jin, Jiayu Gong, and Xiaofei Liao. "Performance and energy modeling for live migration of virtual machines", ACM, In Proceedings of the 20th international symposium on High performance distributed computing, HPDC '11, pages 171‑182, New York, NY, USA, 2011.

[18] Anton Beloglazov, Jemal Abawajy, and Rajkumar Buyya, "Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing", Future Generation Computer Systems, (0):‑, 2011.

[19] Rodrigo N. Calheiros, Rajiv Ranjan, Anton Beloglazov, César A. F. De Rose, and Rajkumar Buyya, "CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Povisioning Algorithms", Software: Practice and Experience (SPE), Volume 41, Number 1, Pages: 23-50, 2011.

TABLE IV.   POWER CONSUMPTION BY THE SELECTED SERVERS AT DIFFERENT LOAD LEVELS IN WATTS

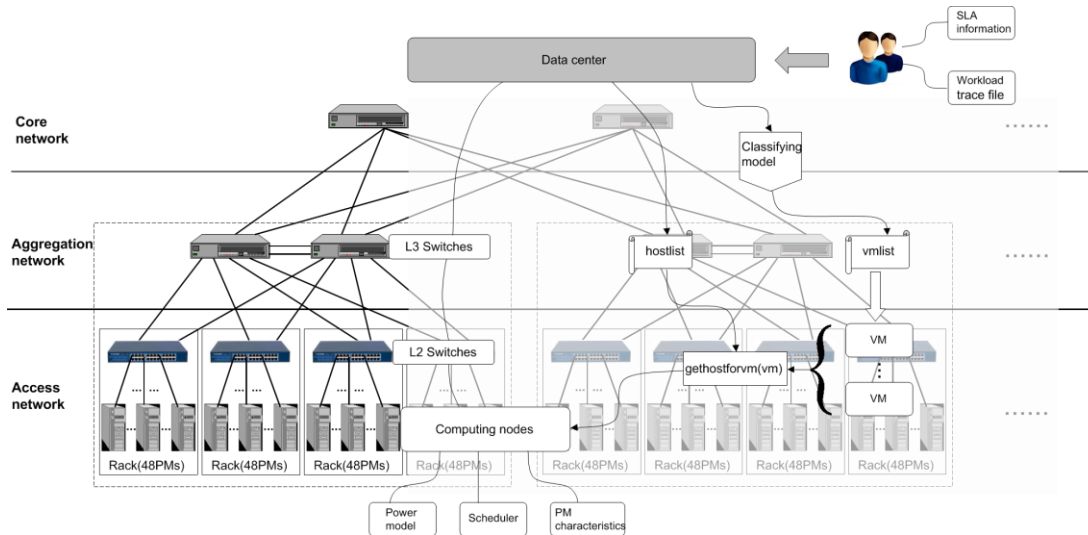| Servers(PMs)	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
HP ProLiant ML110 G4	86	89.4	92.6	96	99.5	102	106	108	112	114	117
HP ProLiant ML110 G5	93.7	97	101	105	110	116	121	125	129	133	135
HP ProLiant DL120 G5	69.5	75.5	84	91.7	101	109	116	121	126	132	136
HP ProLiant DL180 G5	106	117	127	137	146	156	164	172	178	184	189

Figure 7.   The process of the CloudSim simulation.