

A Novel Energy Optimized and Workload Adaptive Modeling for Live Migration

Bing Wei

Abstract—Live migration provides desirable benefits in the field of energy saving, which packs service into fewer physical servers while maintaining the performance level, most of existed work concentrates on the live migration mechanism, however, few work investigates the energy guided live migration decision making, and the varies characteristics of workload always challenges the problem. In this paper we present energy guided and workload adaptive modeling for live migration, two models are developed respectively including energy guided migration model and workload adaptive model, the former model selects the best migrating virtual machine (VM) candidate with the minimal energy consumption while the later model chooses the best migrated physical server candidate in terms of both energy and workload characteristics, furthermore, concerning the service quality, workload adaptive model also takes charge of the determination of the live migration moment. Taking workload characteristics into account in the workload adaptive model avoids considerable unnecessary live migrations and achieves stable live migration with varies workload, and thus reduces the energy usage. The experiments results show that our approach achieves significant energy saving and robust live migration.

Index Terms—Live migration, energy consumption, workload, modeling.

I. INTRODUCTION

Energy saving in data center has received increasing intention recently, virtualization [1], [2] as a powerful tool improves energy efficiency and offers server consolidation and application isolation, many companies' data centers adopt server virtualization to provide flexible and cost-effective resource sharing, for example Amazon EC2 and GoGrid both use Xen virtualization [3] to support multiple virtual machine instances on one single physical server. Live migration as the core technique enables one VM from original server migrating to another server while the VM continues operating, which indicates data center to pack workload into fewer servers to achieve energy saving as well as cost saving. As we know, many existed studies research on live migration strategy, concerning about when and where a VM should be migrated [3]-[5], and existed work about power consumption in data center mostly studies the mechanism of VM placements [3], [6] or the hardware design to eliminate power consumption or demonstration of significant importance of this policy to energy saving[7,8,9],

however, few has focused on the issue when and where a VM should migrate in an energy guided way. Our previous work [10] proposes such an energy aware modeling for live migration, which tells the best migrating VM candidate, the best migrated destination server candidate and the migration moment, however, it would be infeasible for varies workload situations due to the unpredictable influences on the migration energy cost by workload.

Some other related work includes Haikun L's study[11] that models performance and energy consumption during migration to guide cost efficient migration. This study gives a theoretical analysis and empirical researches about data transferred and time consumed in one migration and verifies that the migration energy consumption is proportional to the data transmitted during the migration. Sandpiper [5] proposes both gray-box and black-box strategies to monitor resource usage, detect hotspots and then migrate hotspots by employing live migration, and Tarighi presents a scheduling mechanism for live migration based on fuzzy decision making[12].

In this paper we present an approach to modeling live migration in terms of both energy consumption and various workload, in order to robustly handle the changeable workload and realize minimal energy cost, two models are established respectively: energy guided migration model and workload adaptive model. Energy guided migration model is in charge of selecting the best migrating VM which minimizes the energy usage with no performance disturbing, workload adaptive model gives the best migrated physical server in an energy efficient way, considering the complex workload characteristics, a statics method is adopted in workload adaptive model which uses a large quantity of data for clustering and then generates several clusters to represent workload behaviors, in addition, an optimal cluster which describe the optimal energy efficiency workload is obtained, all these clusters helps to guide the migrating VM to optimal destination host. Furthermore, combining with the performance level measurement, workload adaptive model also tells when a VM is migrated. Due to the analysis of the workload characteristics, our approach is more feasible and effective to handle energy saving for live migration under complex workload circumstance.

The paper is organized as follows: Section 2 introduces the energy guided migration model which selects the migrated VM according to the energy cost. We formulate the energy cost to calculate and evaluate the general power consumption of each VM if it is migrated. Section 3 introduces the other model: workload adaptive model. Based on the theory that each physical server has one optimum workload that minimizes the energy consumption per workload, workload adaptive model gives an optimum solution to destination

Manuscript received March 20, 2012; revised March 29, 2012. This work was supported by the National Grand Fundamental Research 973 Program of China (No. 2010CB328105, No. 2009CB320504), and the National Natural Science Foundation of China (No. 61020106002).

B. Wei is with the Department of Computer Science and Technology, Tsinghua University, Beijing 10084, China (email: weibing@csnet1.cs.tsinghua.edu.cn)

hosts under changeable workload conditions by using a static method. Section 4 and 5 give the implement details and some experimental results which demonstrate the effective of our approach.

II. ENERGY GUIDED MIGRATION MODEL

When one physical server requests live migration, there would be several options to the selection of migrating VM, and clearly comprehensively considering the energy cost and the service quality one option would be optimal. Energy guided migration model would first analyze the key parameters affecting the migration performance and power consumption and then select the best migrating VM candidate.

A. Live Migration Mechanism

Live migration takes a running VM from one physical server to another with no disruptive performance level. It needs several steps to complete:

- 1) Initialize the migration process by selecting the VM and the destination physical server;
- 2) pre-copy the memory of the VM to the destination physical server;
- 3) transfer non-memory state of the VM;
- 4) transfer the control to the destination physical server and re-start the VM at the destination physical server, then remove it from the original physical server.

Most of the energy are consumed in the pre-copy step, during which the CPU and network are intensively working to copy and transfer VM's memory to destination server host, it is a daemon transmitting VM's memory images to destination while still enabling the VM's running, which causes iteratively transmitting memory images generated during previous pre-copy process, the pre-copy terminates on several conditions and we only consider one of them in this paper which is the remaining dirtied image pages are smaller than a predefined value. The size of VM memory is clearly crucial to the power usage during the live migration: the VM with larger memory would need longer migration time and greater usage of power.

Performance would unavoidably decrease during live migration even though VM continues running: the physical resource cannot afford the demand of the workload and even worse that the CPU and network intensive migration process aggravates the confliction of resources. There also would be the downtime during which the VM is unavailable which consists of the time to quiescence the VM on the source PM, transfer the device state to the destination PM, load the device state, and copy over all the remaining memory pages concurrently with loading the device state, however, this downtime is trivial and generally is stable, so mitigating performance degradation should reduce the migration time especially the pre-copy time.

Live migration is a CPU and network intensive process, so low CPU reservation for live migration would impact the migration rate and thus increase the migration time, the CPU available reserved for live migration varies with VM causes different effect on performance, note that network bandwidth available is also reduced due to the background pre-copy,

however, on a gigabit LAN, this impact is small.

Concerning all factors discussed above, which are VM memory, pre-copy time and CPU available reservation for live migration, this selection problem can be modeled specifically in mathematic way.

B. Live Migration Model

Symbols definitions are given as follows: the size of VM memory V_0 , memory dirtying rate D and the network transmission rate R , remaining memory threshold V_{thr} , the dirty-transmit rate $\lambda = D/R$, the times pre-copy needs proceed n .

The total column of data transmitted, corresponding pre-copy time and the pre-copy times are quantitatively expressed as follows using method in [9]:

$$V = V_0 \frac{1 - \lambda^{n+1}}{1 - \lambda} \quad (1)$$

$$T = V / R = \frac{V_0}{R} \frac{1 - \lambda^{n+1}}{1 - \lambda} \quad (2)$$

$$n = \left\lceil \log_{\lambda} \frac{V_0}{V_{thr}} \right\rceil = \log_{\lambda} \frac{V_0}{V_{thr}} + \varepsilon \quad (3)$$

We give a mathematic expression to estimate the migrating cost of each VM, then the VM with smallest cost is selected as the best candidate, which should satisfy the minimal energy consumption and maintain certain performance level, so in order to build this overall expression, some correlations must be investigated: it is verified that energy consumption has a linear correlation with the total data volume transmitted. And to guarantee the service quality, a threshold of migration time represents the longest latency maintaining a certain performance is defined and performance would decrease exponentially when migration time increases compared to the threshold.

Concerning about CPU reservation for live migration, it is reasonable to assume that it has a negative correlation with the migration rate when it is lower than a certain value and it has no influence on the migration rate if it exceeds the value, the available CPU reservation for live migration is calculated by the sum of the remained CPU resource and 30% of the occupied CPU of the VM.

So the final equation is given as follows:

$$C(VM) = \alpha V g(v_{cpu}) \exp\left(\frac{\beta}{1 - \frac{T}{T_{cons}}}\right) \quad (4)$$

$$g(v_{cpu}) = \begin{cases} 1 & v_{cpu} \geq v_{thr} \\ 1 + \gamma \frac{v_{thr} - v_{cpu}}{v_{thr}} & v_{cpu} < v_{thr} \end{cases} \quad (5)$$

where T_{cons} is the threshold of migration time, v_{thr} is the CPU reservation threshold, α , β , γ are coefficient of VM memory, migration time and CPU provision of live migration respectively. $g(v_{cpu})$ denotes the reservation CPU influence factor, when the reservation CPU exceeds the threshold, it's value is 1, if the reservation CPU is below the threshold, it increases the cost.

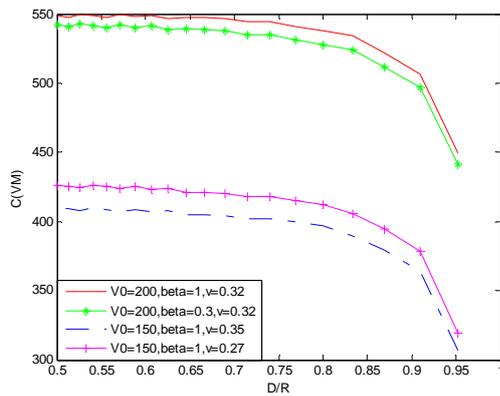


Fig. 1. Cost of VM when $D=20$, R varies from 20 to 40

The cost of VM migration with different parameters is plotted in Fig. 1, clearly, the memory of VM is the greatest influence factor to migration cost, so minimizing the migration cost means selecting the VM with smaller memory and larger CPU reservations.

The energy guided model solves the problem of selecting one VM to be migrated to achieve minimum energy cost. When workload becomes heavy for one physical server, it measures the energy cost function (4) and then gives the one optimal VM.

There would be one case that remove one VM perhaps is not enough for this physical server, in this case it is preferable to remove one VM with heavy workload than two VM with slight workload since we would rather one application suspend than two. So we always first search for one VM that optimizing the power consumption, if failed, we would search for two VMs, the energy of which is calculated in the same way as the energy of one VM is calculated.

III. WORKLOAD ADAPTIVE MODEL

Each physical server exists an optimal workload which minimize the energy consumption per workload, so when live migration request occurs, it is desirable to migrate the VM to one physical server which most optimize the energy efficiency of whole system, however, complex and changeable workload challenges this selection problem and causes dispensable live migrations. For example, the lifetime of the physical server would influence its execute rate, high temperature would impact the work of physical servers, and more importantly the various workload would performs differently in different physical servers. So it is crucial to take workload into account when making decision about the best physical server candidate. So analyzing the workload characteristics of each VM and optimal workload for each physical server, we establish workload adaptive model, it aims to saving energy through selecting the physical server which optimizes the energy efficiency and determining the suitable migration moment to avoid unnecessary migrations as well.

To build such an energy saving model that adaptively varies with the workload while maintaining the service quality, three correlations should be investigated: the correlation between workload and performance, the correlation between workload and migration moment, the correlation between workload and energy consumption.

A. Workload and Performance

It is well known that heavier workload would generally cause severer performance degradation, however, the “heavy” is different for different physical servers and even different configured VMs, some CPU intensive workload would decrease the service quality on this server but would not on another server, considering the specific correlation between workload and performance, several evident factors influence it such as the workload characteristics, the physical attributes of the server, the configuration of VMs, meanwhile, several obscure factors also affect the correlation in non-trivial way such as the operating environment, the lifetime of the server, and even taking into account the conspicuous factors the specific influence is unpredictable, so the mathematic way to build an equation to model this correlation is infeasible and impossible, it would only be possible to build one model under rigorous circumstance of unitary workload and uniform configuration of VMs, however, the characteristics of workload can be obtained by using a statics method, which collects a large quantity of workload data and then clusters into several representative workload clusters, analyze the performance associated with the clusters is simpler.

We focus on three aspects of workload in this paper: CPU utilization, I/O disk utilization and network utilization. The performance level is measured by the respond latency; the greater of the respond latency implies the heavier degradation of the service quality. So the clustering data consists of two parts including the workload part and the performance part, it can be expressed by $[(CPU, I/O, Net), P]$, the workload part of data is used for clustering while the performance part regulates the restriction of service quality.

Several clustering algorithm can be adopted, e. g. hierarchical and k-means clustering, the subtractive clustering is selected due to its speed and robustness. It assumes that each data point is a potential cluster center and calculate the distance of other points to the center, if the distance exceeds a predefined radius, it means this point does not belong to this cluster, the subtractive clustering algorithm selects the best candidate center of one cluster and removes all clustered points then determines the next cluster, which speeds up the clustering process and convergence rate. The clustering would terminate on the condition that all points are within a radius of the corresponding cluster center, the radius represents the deviation to cluster center and it is obvious that smaller radius results in a larger amount of small clusters while contrarily larger radius causes a smaller number of big clusters. The clustering algorithm results in several clusters with one center and one radius, which exemplify characteristics of workload, this clustering algorithm is paralleled executed by each VM of each physical server.

The performance level of one cluster we assume is the average value of the performance part of data points within one cluster, so the correlation between performance and workload can be expressed by the clusters set $[\{C_1, P_1\}, \{C_2, P_2\}, \dots, \{C_n, P_n\}]$, given the current workload, the performance can be measured by the performance of the cluster cover the current workload.

B. Workload and Migrate Moment

When workload increases the confliction of physical resources such as CPU and network would be intense and it would cause inevitably degradation of performance, one effective way mitigating performance degradation is live migration, but another problem arises that is the best moment to request live migration.

One traditional solution is that define a performance threshold and once performance level decrease below the threshold physical server requests migration, it is effective and soon reliefs the tense collisions of resource and thus maintains the service quality, however, it is custom that the workload increases for just a trivial period such as one second and then backs to its normal state, under this circumstance, traditional method would trigger dispensable VM migration, which causes quite large and unnecessary power consumption.

So we can see that the determination of migration moment should involve the characteristic of workload, when workload increases and the performance begins to decrease, the migrating process does not proceed immediately in order to avoid the dispensable migration, only if the increase of workload lasts which indicates the heavy workload is persistent then the migrating process is triggered and energy aware migration model selects one migrated VM. Nevertheless, it is hard to predict the trend of future workload, but fortunately it is reasonable to assume that most of the workload type is within the learned characteristics of workload.

So the following steps are done to determine whether to migrate or not: 1) categorize the current workload into one cluster and then 2) readjust the corresponding cluster, re-calculate the center and performance 3) if the performance P of newly adjusted cluster exceeds the performance threshold, request live migration.

By adding the current workload, the clustering results are adaptively updated. In order to increase accuracy of clustering results and accelerate the clustering rate, we clusters the workload characteristic based on the some new workload not the workload a week ago. So we define a time interval T_{int} and only the data within $[t-T_{int}, t]$ is clustered (t represents the current time). Also an exponential weight $\exp(-\frac{t-T_{int}}{T_{int}})$ is attached to each data according to its generate time, which results in the older data contributes smaller influence on the clustering while newer data contributes bigger influence. The clustering data then is updated as $[(CPU, I/O, Net), P, W]$, in this way, there would be one weight value of each cluster indicating the probability of this workload characteristic, we re-express the clustering results as $[\{C_1, P_1, W_1\}, \{C_2, P_2, W_2\}, \dots, \{C_m, P_m, W_m\}]$, if the weight of the cluster decreases below one threshold, we will delete this cluster, and contrarily if one new cluster is generated and its weight is higher than one threshold, we will add this cluster in. this threshold can be set according to the different situations. In this way, the clusters can fully represent the workload characteristic.

C. Workload and Energy Consumption

For each physical server, it is demonstrated that there exists an optimal workload minimizing the energy usage per workload [13], in this paper considering varies workload characteristics it is more reasonable to assume that workload within a certain range is desirable, to obtain these workload, one energy part should be added into the clustering data, which is re-expressed by $[(CPU, I/O, Net), P, W, E]$, the energy E of the cluster is the weighted mean value of energy part value within the cluster, and then the cluster with minimal energy per workload is selected as optimal cluster, to guarantee the accuracy, some tested data under light workload to heavy workload circumstance is required to initialize the general position of optimal workload.

So according to the optimal cluster of each physical server, if the migrating VM migrates to one physical server, it would be preferable that the added workload approaches to the optimal cluster. The physical server with nearest distance between optimal cluster and added workload would be selected as best candidate.

One another issue would be taken into consideration if the added workload is calculated by simply adding current VM workload and current physical server workload: when decision are made about the migrating destination host, some servers with regularly heavy workload happen to operate in a low workload state then are highly potential being the destination which would result in a dispensable VM migration, or in another scenario, one destination host is selected to be the best candidate of the migrating VM which just operates in low workload, however the migrating VM consists of two workload characteristics and unfortunately the other is much heavier that the selected server could not afford, which would cause performance degradation and even worse another VM live migration, it is essential to avoid such power wasteful VM live migrations, thus it is crucial to record the destination servers' workload characteristics as well as the migrating VMs'.

If a migrating VM with a workload characteristics set $Wl_{VM} : [\{C_1, W_1\}, \{C_2, W_2\}, \dots, \{C_m, W_m\}]$ migrates to a physical server with a workload characteristics set $Wl_{PM} : [\{C_1, W_1\}, \{C_2, W_2\}, \dots, \{C_n, W_n\}]$ and an optimal cluster $\{C_{Opt}\}$, the ensemble possible workload set Wl_{VM-PM} would be $[\{C_1, P_1\}, \{C_2, P_2\}, \dots, \{C_l, P_l\}] \quad 1 < l \leq mn$, which can be calculated by

$$\begin{aligned} \{C_k, W_k\} &= \{C_i, W_i\} \times \{C_j, W_j\} = \{(C_i + C_j), W_i W_j\} \\ \{C_k, W_k\} &\in Wl_{VM-PM}, \{C_i, W_i\} \in Wl_{VM}, \{C_j, W_j\} \in Wl_{PM} \end{aligned} \quad (6)$$

We give a distance measurement equation:

$$D = \sum_{Wl_{VM-PM}} W_i \| C_i - C_{Opt} \| \quad (7)$$

This distance measures a weighted average distance from the optimal cluster, the physical server with minimal distance value is selected, since all possible states are concerned, this method mitigates the VM migration vibration and tends to select one physical server optimizing the energy efficiency.

The workload adaptive model has two contributions: the live migration is requested based on the analysis of the VM's

workload characteristics and the performance constrains, which prevents unnecessary VM migrations and thus saves energy; the destination host is selected in terms of both the workload and the energy, taking into account the workload of migrating VM and migrated host avoids the migration vibration and helps to stably select the physical server according to the optimal cluster.

So we combine the energy guided model with the workload adaptive model to solve the live migration problem together. A large data would be collected to initialize the workload clusters for each physical server. The workload adaptive model would then collect the current workload and update the workload clusters in pre-defined time interval. Once the workload exceeds the workload threshold, it would adopt the energy guided model to select one migrated VM, and then apply the workload adaptive model to select one optimal destination physical server. We give the algorithm below.

Algorithm 1 workload adaptive algorithm

1. Initialize for each physical server:
 - define a performance threshold P_{thr} , clustering a large data set $[(CPU, I/O, Net), P, E]$ for physical server and a large data set $[(CPU, I/O, Net), P]$ for each VM, generate initial clusters of physical server and each VM, initial optimal cluster $\{C_{opt}\}$ is obtained.
 2. On-line phase for each physical server:

Input: current workload of each VM and server, energy usage, respond latency of each VM

 - 2.1 Calculate energy usage per workload E of server and performance P of each VM, form the current server data $\{(CPU, I/O, Net), E\}$ and the current VM data $\{(CPU, I/O, Net), P\}$
 - 2.2 Add $\{(CPU, I/O, Net), E\}$ and $\{(CPU, I/O, Net), P\}$ into clustered data, readjust weight, re-clustering
 - 2.3 Adaptively readjust the optimal cluster
 - 2.4 If P of $\{C, P, W\}$ exceeds P_{thr} , turn to 3, else back to 2
 3. Migrated host selection:
 - 3.1 Request live migration, Proceed energy guided migration model, selecting one migrating VM.
 - 3.2 Calculate distance D for each physical server
 - 3.3 Select the server with minimal D , migrate the VM to this server
-

In addition, considering a long term such as one day, if we record the workload, it is easy to find there exists one general workload curve. For example, the workload may be low in the early morning and in the late midnight, but it is very likely that it increases substantially in the later morning and afternoon. So as this general trend can be predicted, it would save energy a lot if we increase the number of running physical servers when workload significantly increases and decrease the number of running physical servers when workload decreases.

IV. IMPLEMENT

We implement this approach based on Xen 3.4 platform, all necessary workload data of server and VM is collected from Xen domain0 periodically.

To initialize the workload adaptive model, some collected data should be given to form a preliminary model, in our implementation the data is tested and collected with

increasing workload, in addition to reduce the computation of clustering algorithm the data are tracked every 10 seconds and correspondingly the clusters are updated every 10 seconds. The data consists of varies resource utilizations can be obtained from the Xen domain0 since domain-0 is a distinguished VM in Xen which is responsible for I/O processing and scheduling, the resource utilization of each VM can be determined by tracking activity events of domain 0.

In energy guided migration model the memory dirtying rate of different workload can be obtained by tracking dirty statistics using Xen's shadow page mode.

V. EXPERIMENT

Several experiments are conducted to verify the effective energy saving of our approach. Six physical servers each configured with 4 virtual machines are tested with varies kinds of workload, to verify the energy saving efficiency of our model, the randomly select migrating VM and physical server method is implemented, to demonstrate the workload adaptive. We do some experiments that simply migrate VM to physical server according to current workload when migration occurs, and we compare the two methods' results with our approach.

Fig. 2 plots CPU utilization dimension of five basic workload characteristics we adopt in this paper, clearly, we can see that workload 4 and 5 are less stable than workload 1, 2 and 3, especially workload 4 which obviously has two workload behaviors in CPU dimension one is as heavy as 57% CPU utilization and the other is as low as 7% CPU utilization. Five different types of workload is assigned to five physical servers, and by allocating the last server one of five types of workload, five different workload experiment environment are formed.

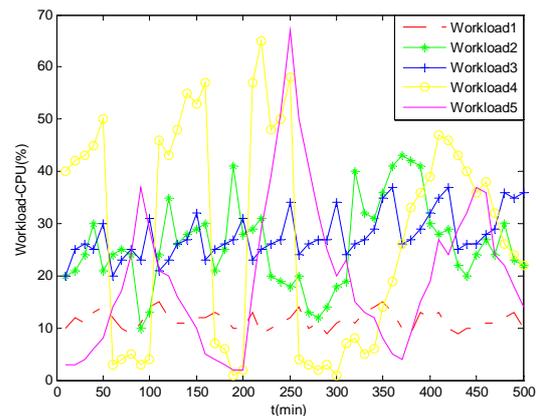
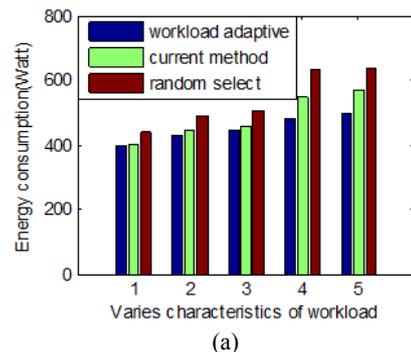


Fig. 2. Five basic workload characteristics



(a)

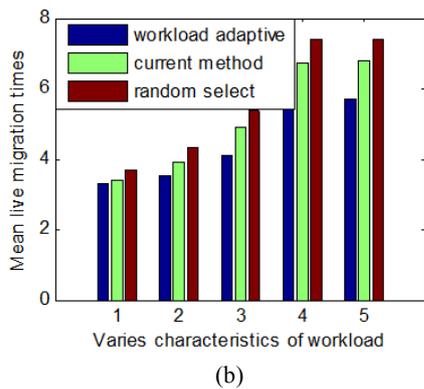


Fig. 3. (a)Energy consumption varies with workload within an hour, (b) average live migration times varies with workload within an hour

We compare the three different methods in two respects: total migration energy cost and average migration time. Fig. 3(a) gives the overall energy consumption within an hour of five different workload with our approach, the current workload method and random select method, and Figure 3(b) illustrates the corresponding average live migration times within an hour, it is clear that our method can achieve significant energy saving than other two methods, and our approach reduces the total migration times, we also can see that the migration times and energy consumption of our method are much smaller when applying workload 4 and 5, it indicates that our method achieves better energy saving when workload is complex.

VI. CONCLUSIONS

In this paper we present a novel workload adaptive and energy optimized modeling for live migration, which consists of energy guided migration model and workload adaptive model. Energy guided migration model selects the best candidate migrating VM when the physical server request live migration. It formulates the energy consumption cost in terms of VM memory and reservation CPU. And workload adaptive model on the other hand choose the best migrating destination server candidate with the minimum power consumption, in addition, it determines the migration moment in terms of both the performance and the trend of workload.

A static method is adopted in workload adaptive model that clusters large quantity of workload data to generate several representative clusters to exemplify workload behaviors. By analyzing workload characteristics considerable unnecessary live migrations are avoided including the unnecessary live migrations when workload would decrease in a moment and migrations when migrates to one physical server whose workload would increase substantially in a few minutes. The selection of migrated server is therefore more stable. Several experiments are conducted which demonstrate that the effective energy saving and workload adaptive of our approach.

Some further work would concentrate on the co-migration scenario where several VMs request live migration simultaneously.

ACKNOWLEDGMENT

This work is supported by the National Grand Fundamental Research 973 Program of China (No. 2010CB328105, No. 2009CB320504), and the National Natural Science Foundation of China (No. 61020106002).

REFERENCES

- [1] J. Lange, K. Pedretti, T. Hudson, P. Dinda, Z. Cui, L. Xia, P. Bridges, A. Gocke, S. Jaconette, M. Levenhagen and R. Brightwell, "Palacios and Kitten: New High Performance Operating Systems for Scalable Virtualized and Native Supercomputing," In *Proceedings of the 24th IEEE International Parallel and Distributed Processing Symposium (IPDPS'10)*, Atlanta, Georgia, USA, April 19-23, 2010, pp.1-12.
- [2] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the Art of Virtualization," In *Proceedings of the nineteenth ACM symposium on Operating Systems Principles (SOSP'03)*, Lake George, New York, USA, October 19-22, 2003, pp.164-177.
- [3] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, A. Ho, R. Neugebauer, I. Pratt, and A. Warfield, "Xen and the art of virtualization," in *Proceedings of SOSP'03*, Oct. 2003.
- [4] K. Sato, H. Sato, S. Matsuoka, "Model-based Optimization for Data-intensive Application on Virtual Cluster," In *The 9th IEEE/ACM International Conference on Grid Computing (Grid'08)*, Tsukuba, Japan, pp.367-368.
- [5] T. Wood, P. Shenoy, A. Venkataramani and M. Yousif, "Blackbox and Gray-box Strategies for Virtual Machine Migration," In *Proceedings of 4th USENIX Symposium on Networked Systems Design and Implementation (NSDI'07)*, Cambridge, MA, USA, April 11-13, 2007, pp. 229-242.
- [6] A. Verma, P. Ahuja, and A. Neogi. pMapper, "Power and Migration Cost Aware Application Placement in Virtualized Systems," In *Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware (Middleware'08)*, Springer-Verlag, Leuven, Belgium, December 1-5, 2008, pp.243-264.
- [7] M. Y. Lim, F. Rawson, T. Bletsch, and V. W. Freeh, "PADD: Power Aware Domain Distribution," In *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems (ICDCS'09)*, Montreal, Quebec, Canada, June 22-26, 2009, pp.239-247.
- [8] M. Cardoso, M. R. Korupolu and A. Singh, "Shares and Utilities Based Power Consolidation in Virtualized Server Environments," In *Proceedings of the 11th IFIP/IEEE international conference on Symposium on Integrated Network Management (IM'09)*, New York, NY, USA, June 1-5, 2009, pp. 327-334.
- [9] L. Hu, H. Jin, X. Liao, X. Xiong and H. Liu, "Magnet: A Novel Scheduling Policy for Power Reduction in Cluster with Virtual Machines," In *Proceeding of 2008 IEEE International Conference on Cluster Computing (Cluster'08)*, September 29- October 1 2008, Tsukuba, Japan, pp.13-22.
- [10] B. Wei, C. Lin, and X. Z. Kong, "Energy Aware Modeling for live migration", 2011.
- [11] H. Liu, C. Z. Xu, H. Jin, J. Gong, and X. Liao, "Performance and Energy Modeling for Live Migration of Virtual Machines," In *HPDC'11*, June 8-11, 2011.
- [12] M. Tarighi, S. A. Motamedi and S. Sharifian. "A New Model for Virtual Machine Migration in Virtualized Cluster Server Based on Fuzzy Decision Making," *Journal of Telecommunications*, vol.1, No.1, February 2010, pp.40-51.
- [13] S. Srikantaiah, A. Kansal, and F. Zhao, Energy Aware Consolidation for Cloud Computing, USENIX HotPower'08, 2008.



Bing Wei received the B. S. degree in computer science and technology (with honors) from Xi'an Jiaotong University, Xi'an, China, in 2009. He is currently working towards the Master degree at Tsinghua University, Beijing, China.

His research interests include the energy saving of live migration, cloud computing.