Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 45 (2015) 823 – 831

Procedia Computer Science

International Conference on Advanced Computing Technologies and Applications (ICACTA-2015)

Migration Performance of Cloud Applications- A Quantitative Analysis

Narander Kumar[a*], Swati Saxena[b]

[a,b]*Department of Computer Science, Babasaheb Bhimrao Ambedkar University (A Central University), Lucknow, India*

Abstract

Performance of a cloud data centre must abide the Service Level Agreement parameters and must provide negotiated Quality of Service values. One of the key areas in cloud computing, where the possibility of performance tuning is the maximum, is live virtual machine migration. Migration entails time and traffic which are crucial factors as far as the QoS is concerned. This paper focuses on quantitative analysis of live migration within a cloud data centre with the aim of understanding the factors which are responsible for cloud's efficiency. Various key parameters, such as, virtual machine size, network bandwidth available and dirty rate of a cloud application are discussed in detail and given the comparisons also, to give a clear view of their role in live migration's performance. The analysis presented in this paper gives a proper platform for considering future enhancements and/or modifications in the existing migration technology.

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).

Keywords: Cloud data-centers; Migration; Performance; Service downtime; CloudSim

1. Introduction

The strength, of virtualization in cloud computing, lies in successful and efficient virtual machine migration. Proper considerations must be given to data centre performance and service availability to the customer before taking the migration decision.

* Corresponding author. Tel.: +91-7500371818
E-mail address: nk_iet@yahoo.co.in

Peer-review under responsibility of scientific committee of International Conference on Advanced Computing Technologies and Applications (ICACTA-2015).
doi:10.1016/j.procs.2015.03.163

Choice of migrating virtual machine, location of the destination host machine, volume of data to be migrated, migration time, network bandwidth available and service downtime are some of the crucial factors that must be analyzed beforehand to ensure an energy-efficient and cost-effective migration. Virtual Machine Migration (VMM) refers to a process of transferring a user's application from one host server to another in a single cloud data centre or within a couple of data centers. A user's application is known as a virtual machine. A physical server may host a number of such virtual machines with varying resource requirements, thereby catering multiple users simultaneously. Reasons for a virtual machine migration can be many-load balancing, energy saving, performance issues, cost efficiency, etc. Migration can be cold, where the running virtual machine is first stopped and then transferred to the destination server. This interrupts the user's work for some time, and is referred as service downtime. Today, most data centers are opting for hot or live migration, where a virtual machine's data is copied to the destination server in an iterative manner while the VM is running at the source server. After some number of copying cycles, the source VM is stopped and remaining data is copied to the destination. This category of migration usually results in less service downtime.

The aim of live VMM must be to reduce the number of iterative copying cycles and service downtime. However, there exists a trade-off between these two factors. This paper discusses the factors which affects the performance of live virtual machine migration, and presents an analytical model of migration performance in terms of these factors. Finally, simulation results are shown to support the proposed model. The paper is organized as follows- Introduction in section 1 is followed by review of previous work in section 2, section 3 introduces the proposed analytical model in detail, section 4 presents the simulation results and discussions, finally section 5 concludes the paper followed by references.

2. Review of Work

Live migration of virtual machine in cloud environment has been studied quite a number of times, analyzing its effects and behavior under different circumstances. First of all, factors that affect the migration latency and resources consumption during migration has been discussed and based on workload characteristics and hypervisor configuration, an alternative technique of migration is proposed that reduces overhead[1]. Next, a migration progress system is studied which predict the progress analytically while maintaining user-defined migration objectives[2]. In order to gain parallelism and achieve higher efficiency, migration of multiple virtual machines is proposed and their affects are recorded[3]. Resource availability is a key issue in migration performance; hence, a prediction based analysis of virtual machine migration is studied where an application's behavior with the availability of different types of resources is correlated[4, 10]. Live migration renders heavy responsibility on cloud data centers whose primary aim is service availability. An analytical performance model of migration is evaluated which show that an effective live migration can reduce service rejection scenarios and total delay[5].

Various parameters that affect live migration of virtual machines and on whom the migration performance depends are discussed[6]. Choice of a virtual machine to migrate depends on many factors like its size, dirty rate, etc. A correct selection of victim virtual machine can not only reduce total migration time but can reduce the service downtime as well. Factors responsible for considering a virtual machine for live migration are dealt with[7]. A step further, migration of virtual machines in clusters is discussed to improve the efficiency of a cloud application[8]. Energy consumption and energy saving are hot topics in cloud computing now-a-days. A combination of energy efficiency along with efficient virtual machine migration is studied[9]. Techniques such as compression and layered copying can greatly reduce migration data, thereby, improving migration time and cloud performance. Effects of such techniques are analyzed[11] and are verified by applying them on Xen virtual machines[19]. A study concentrated on energy-aware heuristics for server consolidation is done which tries to minimize power and maximize utilization of cloud resources by the means of live migration[12]. Various types of failures in cloud data centers are discussed such as network failure, VM failure, etc and their impact on cloud services are studied[13]. An integration of integer programming, queuing theory and control theory is proposed for timing analysis of live virtual machine migration in clouds[14]. An evaluation framework is proposed to help in determining an appropriate PaaS provider to an application with a specific resource requirement[15]. Dominant Resource Fairness Mechanisms are applied to use cloud resources in the best possible manner for different demanding virtual machines[16]. A detailed survey of cloud computing techniques, its challenges, advancements and issues are given, providing a sound base for further research in live virtual machine migration[17]. To improve migration performance, a hierarchical resource management system is proposed which group multiple virtual machines on a multi-core computer[18]. Cases where, dirtying rate of a cloud application is much higher, an optimization of migration technique where the working frequency of virtual CPU is reduced to limit the service downtime is suggested[20]. Resource allocation issue is tackled by proposing an admission control and scheduling mechanism which maintains QoS parameters as specified by the cloud users along with improving profits[21]. Live virtual machine migration introduces additional traffic in the cloud network. In order to reduce total traffic in data centre network, a quadratic assignment problem is considered which tries to balance traffic distribution and reduce congestion hotspots[22]. A balancing virtual machine placement algorithm is proposed aimed at reducing the number of running physical machines, thereby, reducing energy consumption of cloud data centre[23]. An attempt to control network traffic and migration time by presenting an online VM placement algorithm is presented and VM migration scheduling algorithm is proposed for the same[24]. Data similarity over replicated virtual machines is exploited to fasten the migrated VM synchronization process and also to reduce migration data[25].

Most of the reviewed papers try to analyze the performance of live virtual machine migration from performance point of view. However, so far, very limited work has been done to find out the details of factors i.e. volume of data to be migrated, migration time, network bandwidth available and service downtime, on which migration's performance depends and how to optimize the performance these factors in order to get the elicitation of migration statistics. In this paper, we detail out the factors responsible for performance of live migration and suggest ways to tune them irrespective of application type or requirements.

3. Proposed Model

Virtual machine migration means copying the memory content of a VM from source server to a chosen destination server. A live virtual machine migration copies the memory content while the virtual machine is active at the source server. This simply means that while memory pages of a virtual machine are copied from source to destination, some memory pages are being rewritten by the application user at the source, thus initiating a need to copy these newly written pages again to the destination. This results in repeated copying of dirty memory pages from source to the destination. During the first copy cycle, entire operating system memory is copied to the destination. The time taken in the first cycle depends on the size of virtual machine's RAM and the network bandwidth available. During this time, some memory pages will be rewritten, we call them as dirty pages. Number of pages rewritten during the first cycle time depends on the dirty rate of the application. Higher the dirty rate, more number of pages will be rewritten. Second cycle of migration will copy only the pages dirtied during the first cycle of the virtual machine migration. Time taken by the second cycle depends on the number of dirty pages to be copied and the network bandwidth available. As again, during second cycle some pages will be dirtied, thus requiring third cycle which copies pages dirtied during the second cycle. In this manner, there occurs a number of copying cycles in migration. Theoretically, as long as there are dirty pages at the source virtual machine, there will be a round or cycle of copying them to the destination. However, practically, there must be a way to limit the number of copying cycles from source to destination. The condition(s) imposed will not only stop the iterative nature of copying but will also terminate the source virtual machine. This initiates the start of service downtime. During service downtime, all the dirty pages are transferred to the destination in one cycle and the destination virtual machine is synchronized with the source virtual machine. Once the destination VM is synchronized, it starts receiving the requests of the user's application, thereby ending the service downtime period.

Virtual machine migration from source to destination can be considered as consisting of two phases. Beginning of VM migration with the first phase where a limited number of copying cycles will take place. The end of a certain number of copying cycles will mark the beginning of the second phase which is service downtime or service dead phase. As obvious, the end of service dead phase will complete the VM migration procedure.

i.e., Total migration time (T_m) = time taken by copy cycles (T_c) + service downtime (T_{sd}) (1)

Figure 1 given below depicts the two sequential phases of a live VM migration.

Fig.1: Sequential phases of live VM migration

Also the total data migrated during a live virtual machine migration is given as-

Total migrated data (Dm) = data migrated during copy cycles (Dc) + data migrated during service downtime (Dsd) (2)

Let the size of a virtual machine's RAM be V, the dirtying rate of a user's application be R and the network bandwidth available be B. The table given below gives a mathematical formulation of each copy cycle during the first phase.

Table 1: Analysis of phase 1 of live VM migration

No. of copy cycles	Time taken during each copy cycle	Data migrated during each copy cycle	Data dirtied during each copy cycle
1	$\frac{\mathbf{V}}{\mathbf{B}}$	V	$\frac{\mathbf{V}}{\mathbf{B}} * \mathrm{R}$
2	$\frac{V}{B} * \frac{R}{B}$	$\frac{\mathbf{V}}{\mathbf{B}} * \mathrm{R}$	$\frac{V * R^2}{B^2}$
3	$\frac{V * R^2}{B^3}$	$\frac{V * R^2}{B^2}$	$\frac{V * R^3}{B^3}$
---	---	---	---
i	$\frac{V}{B} * \left(\frac{R}{B}\right)^{i-1}$	$V * \left(\frac{R}{B}\right)^{i-1}$	$V * \left(\frac{R}{B}\right)^{i}$

As evident from the above table, data dirtied during the current cycle is the data migrated in the next cycle. Now, to limit the number of copy cycles and mark the beginning of service dead phase, there are some conditions which are also known as 'stop n copy' conditions. For different virtualization environments, there are different 'stop n copy' conditions. For instance, XEN virtualization environment has the following 'stop n copy' conditions-

- A maximum limit on the number of iterations or copy cycles is 29. It means no more than 29 copy cycles can take place during live migration of Xen VMs.
- A maximum limit on the amount of data to be migrated is 3 times the size of RAM, i.e., if more than thrice size of Xen VM RAM is transferred during multiple copying cycles, then the source VM is forcefully stopped and the remaining contents are copied to the destination.
- A maximum limit on the amount of data to be migrated in the next cycle is 50 pages, i.e., if more than 50 pages data needs to be migrated then the VM is stopped and then copying takes place.

Also, for VMware, ‘stop n copy’ conditions are as follows-

- The current copy cycle dirty content can be transmitted in t milliseconds. Here, t is a user setting and is known as switchover goal time.

- Minimum required progress amount be X, i.e., if the difference in the amount of memory dirtied between current and the immediate copy cycle is less than X, then stop condition is activated.
- In addition to the above two conditions, if progress amount is less than X and switchover goal time is more than 100 seconds, VMware migration is considered as a failure.

Now, data migrated during ith copy cycle will always be less than or equal to the size of virtual machine's RAM V, i.e.

$$V * \left(\frac{R}{B}\right)^{i-1} \leq V \quad \text{For i >= 1} \qquad \Rightarrow \left(\frac{R}{B}\right)^{i-1} \leq 1 \tag{3}$$

i.e., dirtying rate will be less than or equal to the network bandwidth available. Also, when R=B, i^{th} cycle mimics the 1^{st} iteration for the time taken and data migrated. When R=B, data dirtied during i^{th} cycle will be equal to the size of VM i.e. V. This suggests that data to be migrated in the $(i+1)^{th}$ cycle is equal to the size of VM and thus this cycle mimics the 1^{st} migration cycle.

Now, total copying time, T_c , up to i^{th} iteration will be-

$$T_c = \frac{V}{B} + \frac{VR}{B^2} + \ldots.. + \frac{V}{B}\left(\frac{R}{B}\right)^{i-1} \Rightarrow T_c = \frac{V}{B}\sum_{k=0}^{i-1}\left(\frac{R}{B}\right)^k \quad \Rightarrow T_c = \frac{V}{B} * \frac{\left(\frac{R}{B}\right)^i - 1}{\frac{R}{B} - 1} \tag{4}$$

2. Total data migrated D_m up to i^{th} iteration will be-

$$D_c = V + \frac{VR}{B} + \frac{VR^2}{B^2} + \ldots\ldots.. + V\left(\frac{R}{B}\right)^{i-1} \Rightarrow D_c = V\sum_{k=0}^{i-1}\left(\frac{R}{B}\right)^k \quad \Rightarrow D_c = V * \frac{\left(\frac{R}{B}\right)^i - 1}{\frac{R}{B} - 1} \tag{5}$$

3. For 'stop n copy' condition to activate in live migration, amount of data to be migrated in the next copy cycle should be less than a particular function of V, i.e., size of virtual machine. Let us call that value as 'copy limit' and is denoted by ϕ. ϕ is a predetermined percentage of V that can be migrated during service downtime, i.e., when

$$V\left(\frac{R}{B}\right)^i \leq \phi V$$

Consider the i^{th} cycle as the last cycle before shutting down the virtual machine completely and copying its content to the destination. If $V\left(\frac{R}{B}\right)^i \leq \phi V$ then service downtime starts which transfers ϕV amount of data to the destination. As obvious, the value of copy limit ϕ will affect the virtual machine downtime and also the no of copy cycle required during a particular live migration. If copy limit ϕ is low, service downtime will be less but number of copy cycles will increase. The value of ϕ is decided by the cloud service provider keeping in mind the type of user's application and the QoS required.

The duration of virtual machine service downtime is $T_{sd} = \frac{\phi V}{B}$ (6)

Also, data migrated during service downtime is $D_{sd} = \phi V$ (7)

Value of ϕ also determines the number of copy cycles required before starting the 'stop n copy' phase, as

$$V\left(\frac{R}{B}\right)^i \leq \phi V \qquad \Rightarrow i \leq \ln_{R/B} \phi \tag{8}$$

Substituting eq[n] 4 and 6 in 1, the total migration time is $T_m = \frac{V}{B}\left[\frac{\left(\frac{R}{B}\right)^i - 1}{\frac{R}{B} - 1}\right] + \phi\frac{V}{B}$

(9)

Also total migration traffic due to one virtual machine's live migration is $D_m = V\left[\frac{\left(\frac{R}{B}\right)^i - 1}{\frac{R}{B} - 1}\right] + \phi V$

(10)

As is quite evident from the mathematical formulations given above, performance of live virtual machine migration depends on the following factors-

- Size of virtual machine V
- Network bandwidth B
- Dirty rate of the application R

4. Simulation Results and Discussion

To analyze the time taken during migration by a typical application and the amount of extra traffic caused by migration in a cloud data centre, we simulated the migration of a virtual machine in a single data centre using CloudSim simulator. We considered a virtual machine with RAM size V of 4GB, a constant network speed B as 60KB/sec and a constant dirty rate R is 2pages/sec. The value of is varied from 5% to 50%. Following figures give the variation in migration time, migration data and service downtime with respect to and also the number of copy cycles. Figure 2 shows the trade-off between copy limit and data migrated during service downtime i.e., Dsd. As is clear from the figure, higher the copy limit, greater amount of migration data will be transferred during service downtime period. This will increase the downtime, resulting in longer service unavailability to the user. An optimum choice of copy limit is, therefore, important to restrict the virtual machine dead phase. Figure 3 compares the copy limit values with the total data transferred during live migration. Again, it shows that limiting the copy limit will increase the migration data during copy cycles.

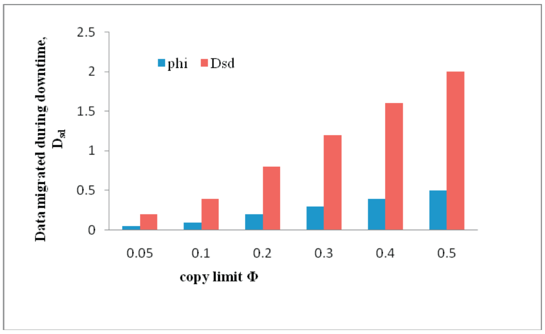

Fig.2- Variation of D_{sd} with copy limit Φ

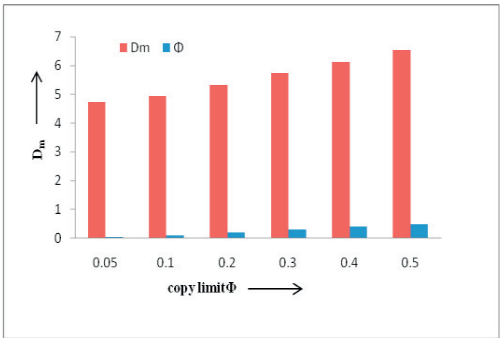

Fig.3- Variation in D_m with copy limit Φ

Figure 4 given below, shows the effect of increasing the number of copy cycles (i) on the data transferred during each cycle (Dc). Figure 4 clearly shows that data migrated in each cycle is dependent on the dirty rate of the application and therefore, increasing the number of cycles gives almost constant data migration statistics. Recall that we have assumed a constant dirty rate of the application.

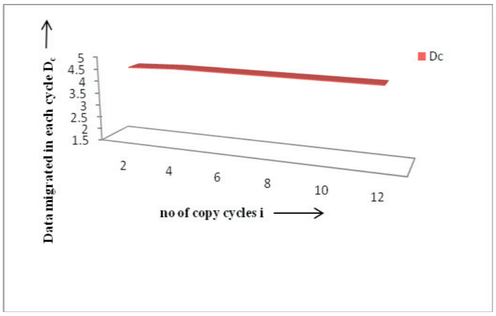

Fig.4- Variation in data migrated in each cycle

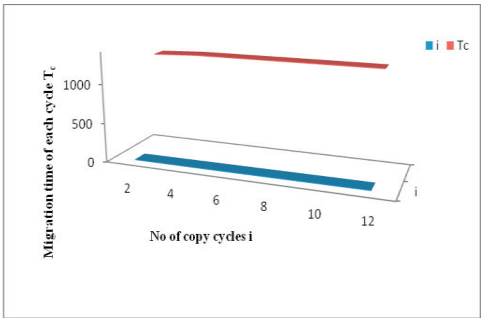

Fig.5- Variation of Migration time with each copy cycle

Figure 5 highlights the variation of copy cycles with migration time in each cycle. As shown, increasing the number of copy cycles does little change in each cycle's migration time. This is due to the fact that, migration time in each cycle depends on the data migrated during the cycle and the network bandwidth available. Figure 6 and 7 compares total migration time and service downtime with copy limit respectively. Total migration time is bound to increase with an increase in the copy limit. It will, however, decrease the number of copy cycles. Service downtime, specially will increase because copy limit declares the amount of data to be migrated during downtime. Consequently, higher the copy limit, higher will be the service downtime. Now, we consider the energy efficiency of live virtual machine data migration by considering the migration duration. The higher the migration time, the longer source physical machine is active; hence the longer it is consuming energy. Our aim in live VM migration should be to minimize energy consumption by reducing the total migration time. As is evident from figure 6, choice of copy limit ϕ plays a very important role in decreasing or increasing the migration duration. Figure 8 shows the variations of energy consumption with respect to copy limit ($T_m(\phi)$) and number of copy cycles i ($T_m(i)$).

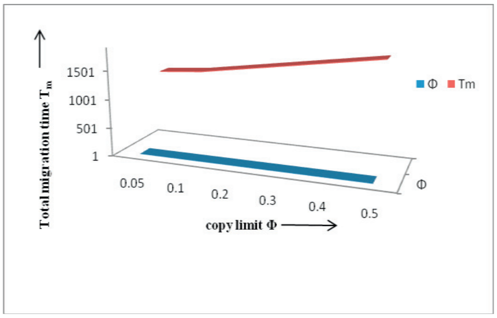

Fig.6- Variation of total migration time with copy limit

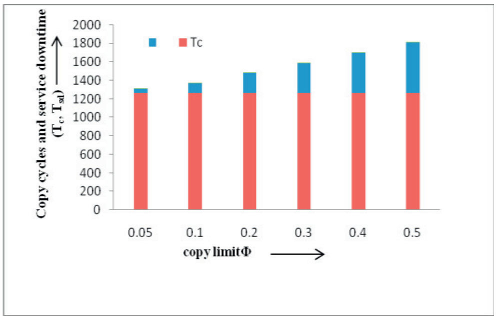

Fig.7- Variation of service downtime with copy limit

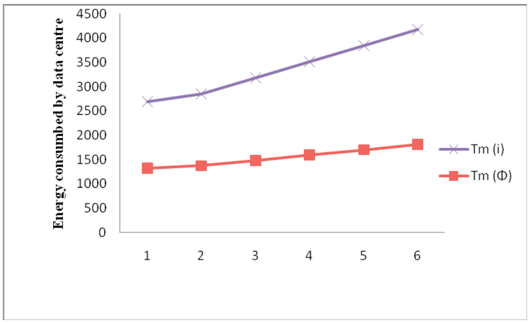

Fig.8- Energy efficiency of VM migration

5. Conclusion and Future Work

This paper presented a quantitative analysis of all factors that affect live virtual machine migration time in shared clouds. As migration is essential to satisfy both service level agreements between the cloud user and the cloud provider and quality of service metrics, focusing on its strengths and weaknesses is crucial. This paper emphasized three factors namely virtual machine size, network speed and dirty rate of the application which play important roles in optimizing the performance of live VM migration. Further, simulation results show the variation of migration load and time with these factors. We would like to emphasize on effective performance of cloud applications by devising an energy-efficient virtual machine placement technique. Also, optimization of migration of multiple machines in a shared cloud needs to be seen as a future work.

References

1. Isci C., Liu J., Abali B., Kephart J.O., Kouloheris J., Improving server utilization using fast virtual machine migration, IBM Journal of Research and Development , 2011:55:4:1-4:12.
2. Jie Zheng, Ng T.S.E., Sripanidkulchai K., Zhaolei Liu, Pacer: A Progress Management System for Live Virtual Machine Migration in Cloud Computing, Network and Service Management, IEEE Transactions on , 2013:10:369-382.
3. Kikuchi, Shinji, Matsumoto, Yasuhide, Using Model Checking to Evaluate Live Migrations, IT Professional , 2013:15:36-41.
4. Yangyang Wu, Ming Zhao, Performance Modeling of Virtual Machine Live Migration, Cloud Computing (CLOUD), 2011 IEEE International Conference on , 4-9 July 2011, p.492-499.
5. Khazaei H., Misic J., Misic V.B., Performance of an IaaS cloud with live migration of virtual machines, Global Communications Conference (GLOBECOM), 2013 IEEE , 9-13 Dec. 2013, p.2289-2293.
6. Akoush S., Sohan R., Rice A, Moore AW., Hopper A, Predicting the Performance of Virtual Machine Migration, Modeling, Analysis & Simulation of Computer and Telecommunication Systems (MASCOTS), 2010 IEEE International Symposium on , 17-19 Aug. 2010, p.37-46.

7. Anala M.R., Kashyap M., Shobha G., Application performance analysis during live migration of virtual machines, Advance Computing Conference (IACC), 2013 IEEE 3rd International , 22-23 Feb. 2013, p.366-372.
8. Kejiang Ye, Xiaohong Jiang, Ran Ma, Fengxi Yan, VC-Migration: Live Migration of Virtual Clusters in the Cloud, Grid Computing (GRID), 2012 ACM/IEEE 13th International Conference on , 20-23 Sept. 2012, p.209-218.
9. Kejiang Ye, Dawei Huang, Xiaohong Jiang, Huajun Chen, Shuang Wu, Virtual Machine Based Energy-Efficient Data Center Architecture for Cloud Computing: A Performance Perspective, Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom) , 18-20 Dec. 2010, p.171-178.
10. Yangyang Wu and Ming Zhao, Performance Modeling of Virtual Machine Live Migration, In Proceedings of the 2011 IEEE 4th International Conference on Cloud Computing (CLOUD '11), IEEE Computer Society, Washington, DC, USA, 2011, p.492-499.
11. Jing Yang, Key Technologies and Optimization for Dynamic Migration of Virtual Machines in Cloud Computing, In Proceedings of the 2012 Second International Conference on Intelligent System Design and Engineering Application (ISDEA '12), IEEE Computer Society, Washington, DC, USA, p.643-647.
12. Zhibo Cao, Shoubin Dong, An energy-aware heuristic framework for virtual machine consolidation in Cloud computing. J. Supercomput. , July 2014, p.429-451.
13. Xiaodong Liu, Weiqin Tong, Xiaoli Zhi, Fu Zhiren, Liao Wenzhao, Performance analysis of cloud computing services considering resources sharing among virtual machines, J. Supercomput. July 2014, p.357-374.
14. Yongqiang Gao, Haibing Guan, Zhengwei Qi, Tao Song, Fei Huan, Liang Liu, Service level agreement based energy-efficient resource management in cloud data centers, Comput. Electr. Eng. July 2014, p.1621-1633.
15. Gültekin Ataş, Vehbi Cagri Gungor, Performance evaluation of cloud computing platforms using statistical methods, Comput. Electr. Eng. July 2014, p.1636-1649.
16. Eric Friedman, Ali Ghodsi, Christos-Alexandros Psomas, Strategyproof allocation of discrete jobs on multiple machines, In Proceedings of the fifteenth ACM conference on Economics and computation (EC '14), ACM, New York, NY, USA, 2014 p.529-546.
17. Frederico Durao, Jose Fernando Carvalho, Anderson Fonseka, Vinicius Cardoso Garcia, A systematic review on cloud computing. J. Supercomput. , June 2014, p.1321-1346.
18. Zengxiang Li, Xiaorong Li, Long Wang, Wentong Cai, Hierarchical resource management for enhancing performance of large-scale simulations on data centers, Proceedings of the 2nd ACM SIGSIM/PADS conference on Principles of advanced discrete simulation (SIGSIM-PADS '14), ACM, New York, NY, USA, 2014, p.187-196.
19. Hai Jin, Li Deng, Song Wu, Xuanhua Shi, Hanhua Chen, Xiaodong Pan, MECOM: Live migration of virtual machines by adaptively compressing memory pages, Future Generation Computer Systems, 2014:38:23-35.
20. Hai Jin, Wei Gao, Song Wu, Xuanhua Shi, Xiaoxin Wu, Fan Zhou, Optimizing the live migration of virtual machine by CPU scheduling, Journal of Network and Computer Applications, 2011:34:1088-1096.
21. Garg Saurabh Kumar, Adel Nadjaran Toosi, Srinivasa K. Gopalaiyengar, Rajkumar Buyya, SLA-based virtual machine management for heterogeneous workloads in a cloud datacenter, Journal of Network and Computer Applications, Available online 1 August 2014.
22. Garg Saurabh Kumar, Adel Nadjaran Toosi, Srinivasa K. Gopalaiyengar, Rajkumar Buyya, SLA-based virtual machine management for heterogeneous workloads in a cloud datacenter, Journal of Network and Computer Applications, Available online 1 August 2014.
23. Xin Li, Zhuzhong Qian, Sanglu Lu, Jie Wu, Energy efficient virtual machine placement algorithm with balanced and improved resource utilization in a data center, Mathematical and Computer Modelling, 2013:58:1222-1235.
24. Shihong Zou, Xitao Wen, Kai Chen, Shan Huang, Yan Chen, Yongqiang Liu, Yong Xia, Chengchen Hu, VirtualKnotter: Online virtual machine shuffling for congestion resolving in virtualized datacenter, Computer Networks, 2014:67:141-153.
25. Balazs Gerofi, Zoltan Vass, Yutaka Ishikawa, Utilizing memory content similarity for improving the performance of highly available virtual machines, Future Generation Computer Systems, 2013:29:1085-1095.