International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 20 (2017) pp. 9285-9289

Virtual Machine (VM) Earlier Failure Prediction Algorithm

Shaima'a Ghazi
Research Scholar, Department of Computer Science,
Jain University, #1/1-1, Atria Towers, Palace Road,
Bangalore, Karnataka, India.
Orcid Id: 0000-0003-3632-6582

Dr.J.Meenakumari
Professor, Department of Computer Science and Engineering,
Alliance College of Engineering and Design,
Alliance University, Chikkahagade Cross, Chandapura-Anekal Rd, Anekal,
Bangalore, India.

Abstract

Reliability of VMs has always been a challenge in a cloud environment. A fault tolerance (FT) framework that performs environmental monitoring, event logging, parallel job monitoring and resource monitoring to analyze the virtual machine reliability and to perform fault tolerance service are very much required to handle these challenges. As a part of fault tolerance mechanism there is a thorough necessity for providing preventive solutions to have continuity of services. Hence the proactive failure prediction of Virtual Machine (VMs) needs to be focused and also to be improved. It is mainly required to reduce the down time and cope up the scalability issues. This paper deals with one such predictive algorithm to enhance the efficiency.

Keywords: Cloud environment, virtual machines, fault tolerance predictive algorithms and proactive.

INTRODUCTION

Cloud computing is a promising technology, which is capable of modifying the ways of computing and accessing storage in the near future [1]. From a business perspective, clouds offer flexible platforms to both cloud providers and application owners. The adoption and deployment of cloud computing platforms have many attractive benefits, such as reliability, quality of service (QoS) and robustness [2]. It conveys the infrastructure, platform and software (applications) as services accessible to customers in a pay-as-you-go model [3]. In the IT industry, these services are referred to as Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS), respectively [4]. The QoS requirements of these services can be defined in terms of a variety of metrics and are formalized in the service level agreements (SLAs) [5].

The processing units in cloud environments are called virtual machines (VMs) which are used to build the infrastructure of a cloud by interconnecting large-scale virtualized data centers, and resources utilized for computing are delivered to the user over the internet in the form of an on-demand service [6]. These VMs keep running in parallel and should execute the tasks as early as could be expected under the circumstances [7].

In the past, failure prediction approach dealt with VM failures in a proactive way rather than waiting for failures to occur and then react to them. However, the proactive approach requires a failure to be predictable. Only after a failure is predicted, a decision is made to do a migration from a deteriorating node to a spare node (this part shall be discussed in another paper). Therefore, a mechanism has to be formulated for handling failures in cloud systems and improving reliability, availability and serviceability.

LITERATURE REVIEW

In VM load balancing mechanism, resource management is accompanied by several advantages that include scalability, QoS, reduced overhead and increased throughput [11]. Resources are generally divided into physical and logical groups. Logical resources provide temporary control over physical resources [12]. Additionally, logical resources support the development of applications and effective communication protocols [13]. Load balancing mechanism, in addition to being considered as one of the methods to increase the fault tolerance (FT) of cloud computing, provides logical resource management in cloud computing [14]. If the physical facilities in cloud computing are placed in a distributed manner, the network resource management develops and improves by using load balancing techniques, thereby increasing FT at the same time [15].

This prompts issue in the scheduling of customer tasks and achieve FT within the available resources that perform an important role in cloud computing [16]. The mechanism of task scheduling not only satisfies users, but also increases utilization of resources [17].

Jialei Liu et al. [5] suggested an initial virtual cluster allocation algorithm to reduce the total consumption of network resource and energy for the VMs in the data center. A proactive coordinated fault tolerance (PCFT) approach adopts a VM coordinated mechanism to anticipate a deteriorating physical machine (PM) in a cloud data center, and then automatically enables the migration of VMs from the deteriorating PM to the optimal target PMs by using particle swarm optimization (PSO) algorithm. This was a very challenging problem, considering its efficiency, effectiveness and scalability requirements. The above problem was solved through a two-step approach, in which a CPU temperature model was first proposed to anticipate a deteriorating PM, and then the optimal target PMs were searched by using an efficient heuristic algorithm. Finally, the performance of the PCFT approach was evaluated by comparing it with five related approaches in terms of the overall transmission overhead, overall network resource consumption and total execution time, while executing a set of parallel applications. (Only one parameter CPU)

Zibin Zheng et al. [19] proposed a component-ranking framework for fault-tolerant cloud applications. In this FT cloud approach, the significant value of a component is determined by the number of components that invoke this component, the significance values of these components, and how often the current component is invoked by other components. After finding out the significant components, the framework automatically proposes an optimal FT selection algorithm to provide optimal FT strategies to the significant components, based on the user-constraints. The experimental results show that the FT cloud approach significantly outperforms other baseline approaches.

Cloud computing exploits virtualization to minimize power consumption, maximize resource utilization and minimize VM transfer time costs in cloud data centers. The key mechanisms in virtualization that provide flexible resource utilization are allocation of resources dynamically to virtual machines and migrating VMs to other PM. Fahimeh Ramezani et al. [20] presented a multi-objective optimization model using the fuzzy PSO, which improves the efficiency of convectional PSO by using fuzzy logic systems and is relied upon to solve the optimization problem. The model is implemented in a cloud simulator to investigate the performance

Alboaneen et al. [21] used glowworm swarm optimization (GSO) algorithm to solve the VM placement problem and to minimize energy consumption and SLA violation. The process of placing VMs on physical hosts is called virtual machine placement (VMP). VMP problems can be either for a fresh VM placement, where a new VM is placed on physical host or for a VM replacement which is the optimization of the existing placement of VMs. After a period of time, violations of SLA may occur due to factors, such as high CPU utilization or high memory usage of the physical host. Hence, some VMs need to be migrated to avoid overutilization that causes degradation of VM performance.

From the literature review observations, the cloud data centers' characteristics are studied on the basis of VM selection policies and migration algorithms. The available and proposed algorithms found that earlier prediction of failure of VM is an important character on which resource utilization is based. In most of the existing algorithms, a single level policy is taken for VM selection and the parameters taken for proactive FT are not considered too.

PROBLEM STATEMENT

From the literature, it is understood that although prediction of VM failure was proactive, it was inefficient, hence in this study, we propose to predict failure of VM much earlier to optimize the utilization of resources. The threshold is dynamically calculated using a training set taking into account five parameters, is fair enough to keep VM resource usage below the maximum threshold and give more accurate results in predicting the failure. The algorithm aims to give earlier failure prediction time based on a real-time physical data point attribute. By predicting the failure of VMs early could result in saving of the resources and reduction of the running cost of the system.

OBJECTIVE

The purpose of the VM earlier failure prediction algorithm is to find the earlier failure prediction time and calculate the threshold on the basis of real-time physical data point attributes. The failure time is calculated in the training set with the help of resources, such as CPU utilization, CPU usage, bandwidth, temperature and memory.

WORK

In the training set, five iterations were done by using 100 VMs in all iterations. The failed VMs were detected with the help of the five resources mentioned above and dynamically calculated the average thresholds of the VM. This algorithm calculates the maximum and minimum threshold of the failure VMs as well as the earlier prediction time.

The input is a set of VM with resources.

The output is prediction time of the failed VM and the threshold value.

The failure points are identified in the training section by running different iterations with the VM and also by calculating the threshold value of each VM.

Data points that occur before each failure points are selected. Then, the failure points are detected in all iterations.

The data points selected in step 2 are grouped in relation to the failure points collected in step 1. These failure groups are assigned to each of the data points and used as failure groups.

The mean and variance are calculated by using normal distribution models for each attribute (CPU utilization, CPU temperature, memory utilization, CPU usage and bandwidth utilization) in one failure group in each cluster and by using the results from these models.

Using the mean and variance, the probability of the target real-time data attributes were calculated for each failure group of VM.

After obtaining the probability of all the real-time attributes in accordance with each failure group, provided that real-time data is given a decision, is used to calculate the earlier prediction time of each failure VM.

The real-time attribute prediction time as well as the lower threshold value and the upper threshold attribute prediction time is calculated for each resource. The upper threshold value and the lower threshold value for the resources may vary. Based on the difference between the upper threshold value and the lower threshold value, the earlier failure prediction time is calculated. Here, the VM with upper threshold value is found as the VM which has attempted the failure.

RESULTS

In order to reduce the impact of failure during the running of an application on the cloud, there should be a mechanism to anticipate the failures as early as possible so that failures can be proactively addressed. FT is an important issue in job scheduling on cloud data centers. The proposed earlier failure prediction method leads to reduced wastage of resources.

The upper and lower thresholds for each resource of the 100 VMs can be seen in the figures (Figures [1-5]). Using the threshold, the VMs that are going to fail and that which are running under the same environment can be predicted earlier. This part will be discussed in detail in the next paper by using a testing set.

The value of the resources, such as CPU utilization, CPU usage, bandwidth, memory and temperature, used in this algorithm is mentioned below in the graph. Additionally, the resource utilization graph for different resources with five different servers is given.

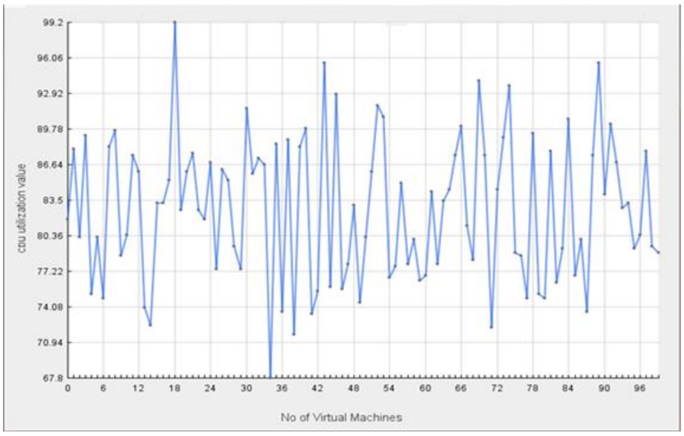

Figure 1: CPU Utilization Graph

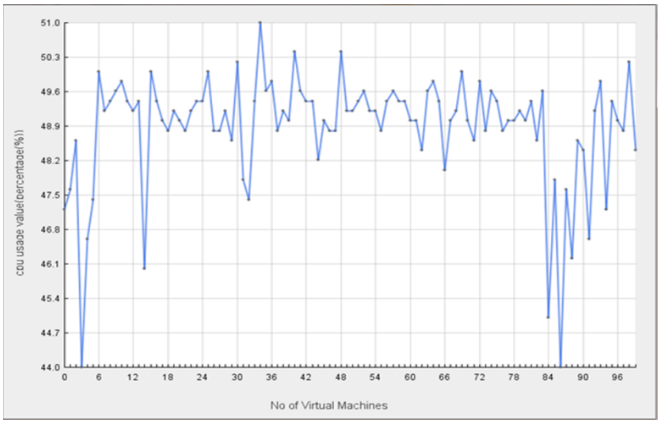

Figure 2: CPU Usage Graph

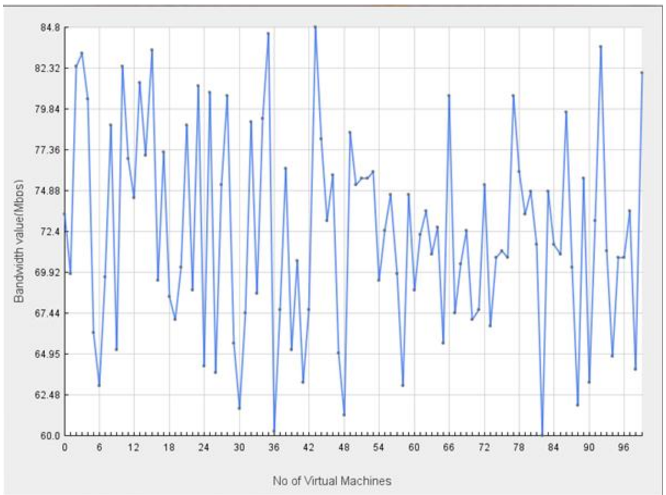

Figure 3: Bandwidth Value Graph

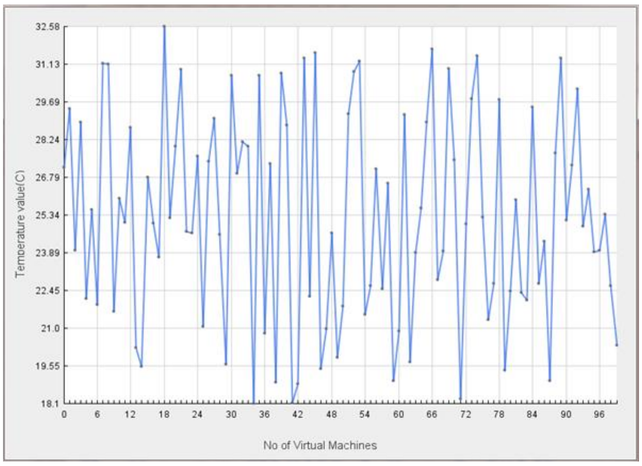

Figure 4: Temperature Value Graph

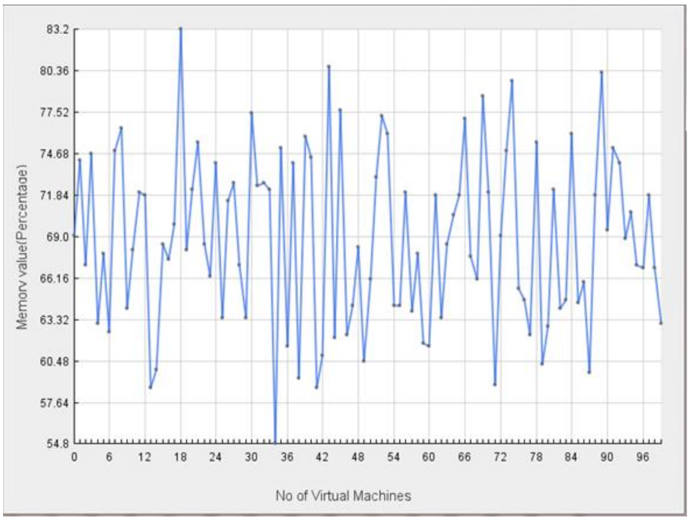

Figure 5: Memory Value Graph

CONCLUSION

Predicting the failure of VMs at an early stage is essential, else it will result in serious problems like wastage of resources, consumption of energy, and rising costs. Proactive FT methods are suitable for failure predictions and thus enable the migration or replacement of VMs, even at the run time of the PM, without affecting the performance of the system.

Finally, the objective of the FT framework is to reduce the wastage of resources and predict failure earlier, thereby minimizing costs, saving time and without violating the SLAs.

LIST OF ABBREVIATIONS

FT	Fault Tolerance
VM	Virtual Machine
QoS	Quality of Service
IaaS	Infrastructre as a Service
PaaS	Process as a Service
SaaS	Software as a Service
SLA	Service Level Agreement
PCFT	Proactive Coordinated Fault Tolerance
PM	Physical Machine
PSO	Particle Swarm Optimization
GSO	Glowworm Swarm Optimization
VMP	Virtual Machine Placement

REFERENCES

[1] Bansal, N., Awasthi, A. and Bansal, S., 2016, “Task scheduling algorithms with multiple factor in cloud computing environment,” In Information Systems Design and Intelligent Applications (pp. 619-627), Springer, New Delhi.

[2] Cheraghlou, M.N., Khadem-Zadeh, A. and Haghparast, M., 2016, “A survey of fault tolerance architecture in cloud computing,” Journal of Network and Computer Applications, 61, pp. 81-92.

[3] Ahmad, R.W., Gani, A., Hamid, S.H.A., Shiraz, M., Xia, F. and Madani, S.A., 2015, “Virtual machine migration in cloud data centers: a review, taxonomy, and open research issues,” Journal of Supercomputing, 71(7), pp. 2473-2515.

[4] Zhu, X., Wang, J., Guo, H., Zhu, D., Yang, L.T. and Liu, L., 2016, “Fault-tolerant scheduling for real-time scientific workflows with elastic resource provisioning in virtualized clouds,” IEEE Transactions on Parallel and Distributed Systems, 27(12), pp. 3501-3517.

[5] Liu, J., Wang, S., Zhou, A., Kumar, S., Yang, F. and Buyya, R., 2016, “Using proactive fault-tolerance approach to enhance cloud service reliability,” IEEE Transactions on Cloud Computing. pp.1-5.

[6] Bruneo, D., 2014, “A stochastic model to investigate data center performance and QoS in IaaS cloud computing systems,” IEEE Transactions on Parallel and Distributed Systems, 25(3), pp. 560-569.

[7] Beloglazov, A., Abawajy, J. and Buyya, R., 2012, “Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing,” Future Generation Computer Systems, 28(5), pp. 755-768.

[8] Gao, Y., Guan, H., Qi, Z., Hou, Y. and Liu, L., 2013, “A multi-objective ant colony system algorithm for virtual machine placement in cloud

computing," Journal of Computer and System Sciences, 79(8), pp. 1230-1242.

[9] LD, D.B. and Krishna, P.V., 2013, "Honey bee behavior inspired load balancing of tasks in cloud computing environments," Applied Soft Computing, 13(5), pp. 2292-2303.

[10] Garg, S.K., Versteeg, S. and Buyya, R., 2013, "A framework for ranking of cloud computing services," Future Generation Computer Systems, 29(4), pp.1012-1023.

[11] Tordsson, J., Montero, R.S., Moreno-Vozmediano, R. and Llorente, I.M., 2012, "Cloud brokering mechanisms for optimized placement of virtual machines across multiple providers," Future Generation Computer Systems, 28(2), pp.358-367.

[12] Beloglazov, A. and Buyya, R., 2013, "Managing overloaded hosts for dynamic consolidation of virtual machines in cloud data centers under quality of service constraints," IEEE Transactions on Parallel and Distributed Systems, 24(7), pp.1366-1379.

[13] Liu, H., Jin, H., Xu, C.Z. and Liao, X., 2013, "Performance and energy modeling for live migration of virtual machines," Cluster computing, 16(2), pp.249-264.

[14] Zissis, D. and Lekkas, D., 2012, "Addressing cloud computing security issues," Future Generation computer systems, 28(3), pp.583-592.

[15] Zhang, Q., Cheng, L. and Boutaba, R., 2010, "Cloud computing: state-of-the-art and research challenges," Journal of internet services and applications, 1(1), pp.7-18.

[16] Warneke, D. and Kao, O., 2011, "Exploiting dynamic resource allocation for efficient parallel data processing in the cloud," IEEE transactions on parallel and distributed systems, 22(6), pp.985-997.

[17] Jhawar, R., Piuri, V. and Santambrogio, M., 2013, "Fault tolerance management in cloud computing: A system-level perspective," IEEE Systems Journal, 7(2), pp.288-297.

[18] Xiao, Z., Song, W. and Chen, Q., 2013, "Dynamic resource allocation using virtual machines for cloud computing environment," IEEE transactions on parallel and distributed systems, 24(6), pp.1107-1117.

[19] Zheng, Z., Zhou, T.C., Lyu, M.R. and King, I., 2010, "FTCloud: A component ranking framework for fault-tolerant cloud applications," In Software Reliability Engineering (ISSRE), 2010 IEEE 21st International Symposium on (pp. 398-407). IEEE.

[20] Ramezani, F., Naderpour, M. and Lu, J., 2016, "A multi-objective optimization model for virtual machine mapping in cloud data centres," In Fuzzy Systems (FUZZ-IEEE), 2016 IEEE International Conference on (pp. 1259-1265). IEEE.

[21] Alboaneen, D.A., Tianfield, H. and Zhang, Y., 2016, "Glowworm Swarm Optimisation Algorithm for Virtual Machine Placement in Cloud Computing," In Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences (pp. 808-814). IEEE.