A Dynamic Resource Management with Energy Saving Mechanism for Supporting Cloud Computing

Liang-Teh Lee, Kang-Yuan Liu, Hui-Yang Huang and Chia-Ying Tseng

Department of Computer Science and Engineering, Tatung University, Taiwan
ltlee@ttu.edu.tw

Abstract

Cloud computing is becoming an important opportunities for industry to provide a high degree of scalability and serviceability of computing resources. In order to integrate the system resource, utilize the resource flexibly, save the energy consumption, and meet the requirements of users in the cloud computing environment, one of the positive solutions is to apply the virtualization technology. A dynamic resource management with energy saving mechanism has been proposed in this paper which presents a method of dynamic voltage scaling for dynamic adjustment of resources by inspecting CPU utilization in the cloud computing environment. The voltage of the idle or light loaded computer can be reduced and heavy loaded works can be migrated to those machines with lighter loading for achieving the purpose of energy saving. The experimental results show that at the user unaware situation, energy consumption can be saved significantly by applying the proposed mechanism.

Keywords: *Cloud computing, virtual machine, dynamic voltage scaling*

1. Introduction

Cloud computing possesses the availability and the scalability, but due to the multi-level, distributed, and virtualized systems, the efficiency of energy be faced with more challenges [1, 2]. Thus, how to save the energy of the data center without affecting both the economic efficiency and system performance is an important issue.

In the cloud computing environment, the amount of workload will affect the loading of the physical machine of cloud server. Virtualization technology can simulate a variety of different platforms and manage the resources of the system. By applying the virtualization technology, in accordance with the requirements of users to configure a virtual machine, both the computing environment and resource management problems can be solved [3]. If the appropriate voltage regulation can be applied, the energy efficiency and the system overhead would be improved [4]. This paper proposes a mechanism to adjust the system voltage based on the CPU utilization, and migrating tasks in a heavy loaded machine to idle machines, so as to improve the resource utilization and reduce the energy consumption.

2. Related Work

2.1 Cloud Computing

The concept of cloud computing [5] represents a dynamic allocation of resources. Through a cluster of multiple servers, could providers can provide services to users. In the could computing environment, the virtualization technology can be applied to allocate resources to achieve the purpose of the dynamic adjusting of resources. Cloud computing is a computing environment based on the Internet and can be extended dynamically. It uses “As a service” network technology to provide users with a large number of service nodes. Providers who provide cloud services, rapid deployment of resources based on virtualization technology, can provide users with a variety of services. Furthermore, users can obtain resources and a variety of services according to their needs. Therefore, regardless of the computing power of software resources or storage capacity of network, users can obtain those resources through the cloud services supplier, and significantly reduce the cost of the software and hardware purchasing.

2.2 Virtualization Technology

Virtualization technology [6, 7] can be classified into two categories, full virtualization and para-virtualization. The advantage of full virtualization is that it can maintain consistent compatibility regardless of hardware environments, however, will increase a greater system loading of the physical machine relatively. While the advantage of para-virtualization is that it may share hardware resources with the original operating system, but the system kernel is necessary to be modified. In terms of performance, the para-virtualized is better than that of the full virtualization, but the corresponding hardware support is insufficient.

Virtualization architecture consists of three components, hardware, Virtual Machine Monitor (VMM), and Virtual Machine (VM). VMM is one of the core implementation of virtualization, and is responsible for providing virtualization and managing hardware resources. There are many virtual machines exist, such as Xen and KVM [8].

2.3. Dynamic Voltage Frequency Scaling (DVFS)

Dynamic voltage frequency scaling is a hardware technology that can dynamically adjust the voltage and frequency of the processor in execution time. By applying DVFS technology, without having to restart the power supply, system voltage and frequency can be adjusted in accordance with the specification of the original CPU design into a different working voltage. While CPU works in lower voltage, the energy consumption can effectively be saved.

The power consumption of the CPU is measured by multiplying the voltage square with the system frequency (as Eq. 2.1). Where V is the voltage, F is the frequency, and C is the capacitive load of the system. The DVFS is the power saving technology by reducing the voltage supply [9]. The reduction of CPU frequency means that the voltage can also be dropped, though it will result in the degradation of the system performance and lead to prolong the execution time. In addition, the overhead of the voltage adjusting should also be considered.

The purpose of the DVFS is to allow the degradation of the execution speed of a task by decreasing the CPU frequency and voltage to reduce the power consumption. This technology is often used in real-time systems.

$$P = V^2 * F * C \tag{2.1}$$

As mentioned above, by applying the DVFS technology, CPU voltage can be lowered, but the execution speed of the task will be reduced. From Eq. 2.1, we can see that if only reducing the frequency, energy cannot be saved effectively. In the system, C is the capacitive load of the system, only in lowering the frequency and also reducing the voltage, the power consumption can be saved effectively.

2.4. CloudSim

In this paper, CloudSim [10] is used to evaluate the performance of the proposed mechanism. It is a cloud computing simulation software and developed by the University of Melbourne. CloudSim can support the construction of the large-scale cloud computing environment for simulation and the data center, service agent, scheduling, and resource allocation platform can also be constructed automatically. Moreover, it provides a virtualization engine that can be used to establish and manage a variety of virtualization services in the nodes of the data center.

While allocating processors for virtualization services, resources sharing in temporal and spatial can be switched flexibly. CloudSim applies mature virtualization technology, according to the requirements of users; the data center can allocate appropriate resources to each user. CloudSim provides resource management and monitoring as well as the mapping of the host and virtual machines. In addition, it provides the function for message exchange in data center and users can develop their own algorithms to be processed in CloudSim. Figure 1 shows the schematic diagram of CloudSim.

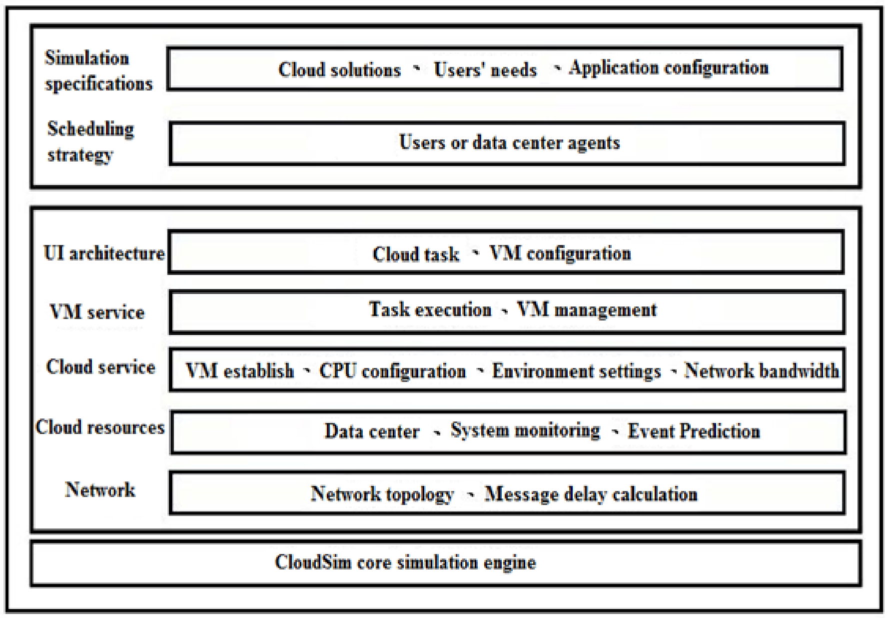

Figure 1. Schematic Diagram of CloudSim

2.5 Virtual Machine Migration

Virtual machine migration technique allows a virtual machine running on a physical machine to be migrated to another physical machine. Migration can be classified into two types, offline migration and real-time migration. For offline migration, the current user's state must be suspended or shut down before the migration can be performed, and users will not be able to take any action. For real-time migration, in contrast, it is not necessary to shut down the original virtual machine and the task can be migrated at the user unaware situation. The advantages of the real-time migration include load balancing, power efficiency, and convenient maintenance [11].

Migration of memory is one of the most important issues in real-time migration [12, 13]. Virtual migration consists of the following steps: pre-migration, reservation, iterative pre-copy, stop and copy, as well as the commitment activation. Where pre-copy is the existing virtual machine migration technology. For real-time migration, by applying the pre-copy technique, all memory pages are migrated before VM migration, once the Writable Work Sequence (WWS) is small enough or the predefined iteration numbers reached, the virtual machine will start to be migrated, CPU status and memory pages are transferred to the destination machine. Although pre-copy can compromise the system downtime and total migration time, it cannot guarantee to transfer data continuously if the number of work queue limit is reached, especially when the virtual machine is running read-intensive operations.

3. Dynamic Virtual Resource Management for Energy Saving

The resource utilization improvement and energy saving are important issues in the cloud computing environment. In addition, the strategy to satisfy the quality of service requirements of users and the task scheduling in the multitasking environment should also be considered.

According to a report by Symantec in 2009, about 97% of the respondents pointed out that the company has been discussed the Green IT issues, while 45% have been importing the green energy plan. According to the utilization of server resources in the current market, about 80% of them are idle and the data center is between 20-30%. In general, even in the idle state, 60% of the power is consumed by the cloud servers. The power consumption coupled with cooling equipment, energy consumption will result in 50-100% of the waste of energy. Thus, it is an important issue to provide a resource management mechanism for energy saving.

3.1 System Overview

A mechanism by applying CloudSim to simulate a large scale cloud data center for energy saving has been proposed in this paper. The system consists of three parts: the CPU utilization monitoring, DVFS adjustment, and real-time migration. CPU utilization on each host is monitored in the system. According to the measured CPU utilization, an appropriate process will be performed for saving energy consumption.

In CloudSim, MIPS (million instructions per second) is used to present the capacity of the host machine, the capacity of VM, and the workload requested by the user. Each workload will be distributed to VMs on different hosts. VM_{MIPS} is the amount of MIPS required for the VM and $Host_{MIPS}$ presents the amount of MIPS the host can support. The utilization of the virtual CPU of a virtual machine, V_{cpu}, can be calculated by Eq. 3.1, and the average utilization of CPUs can be calculated by Eq. 3.2.

$$\mathrm{Vcpu}_i = \frac{\mathrm{VM}_{\mathrm{MIPS}_i}}{\mathrm{HOST}_{\mathrm{MIPS}_i}} \tag{3.1}$$

$$\mathrm{CPU}_{\mathrm{utilization}} = \frac{\sum_{i=1}^{n} \mathrm{Vcpu}_i}{n} \tag{3.2}$$

In the simulation, there are 10 hosts and 30 VMs. Each host supports 3 VMs, total capacities of hosts are 600MIPS, and that of the VM are 250MIPS. The architecture is considered to reduce the waste of idle resources of each host, and to ensure 80% resources is available for VMs. On the same host, 3 VMs share the resources of the host. Once the workload of three VMs on a host becomes heavy, the real-time migration is proceeded to ensure each VM can obtain the required resources.

3.2 System Process

As mention before, the measured CPU utilization can be used for determining voltage adjustment. When the utilization is too low, the voltage will be lowered to reduce the power consumption. In contrast, the utilization is too high, in order to maintain user's quality of service, real-time migration of virtual machines will be preceded. CPU utilization will be measured and recorded per unit-time, and maintain the record for a certain period of time after lowered or raised the voltage gradually. By applying the proposed strategy, the measured CPU utilization can be used to decide whether the migration is required.

Initially, the number of adjustments, threshold of the average CPU utilization, maximum number of VMs each host can support, and initializing parameters are set. The CPU utilization and current voltage are monitored continuously. The average CPU utilization is calculated every three time units, once the average CPU utilization is less than 30%, the number of acting VMs on the host is the maximum, and the voltage is greater than the minimum operating voltage, then DVFS is applied to lower the voltage, otherwise the VM migration is performed for merging VMs. If the average CPU utilization is greater than 80% and the voltage is less than the rated voltage, then DVFS is applied to raise the voltage, otherwise, the real time migration is preceded except the only acting VM on the host. The process of the system is shown in Figure 2.

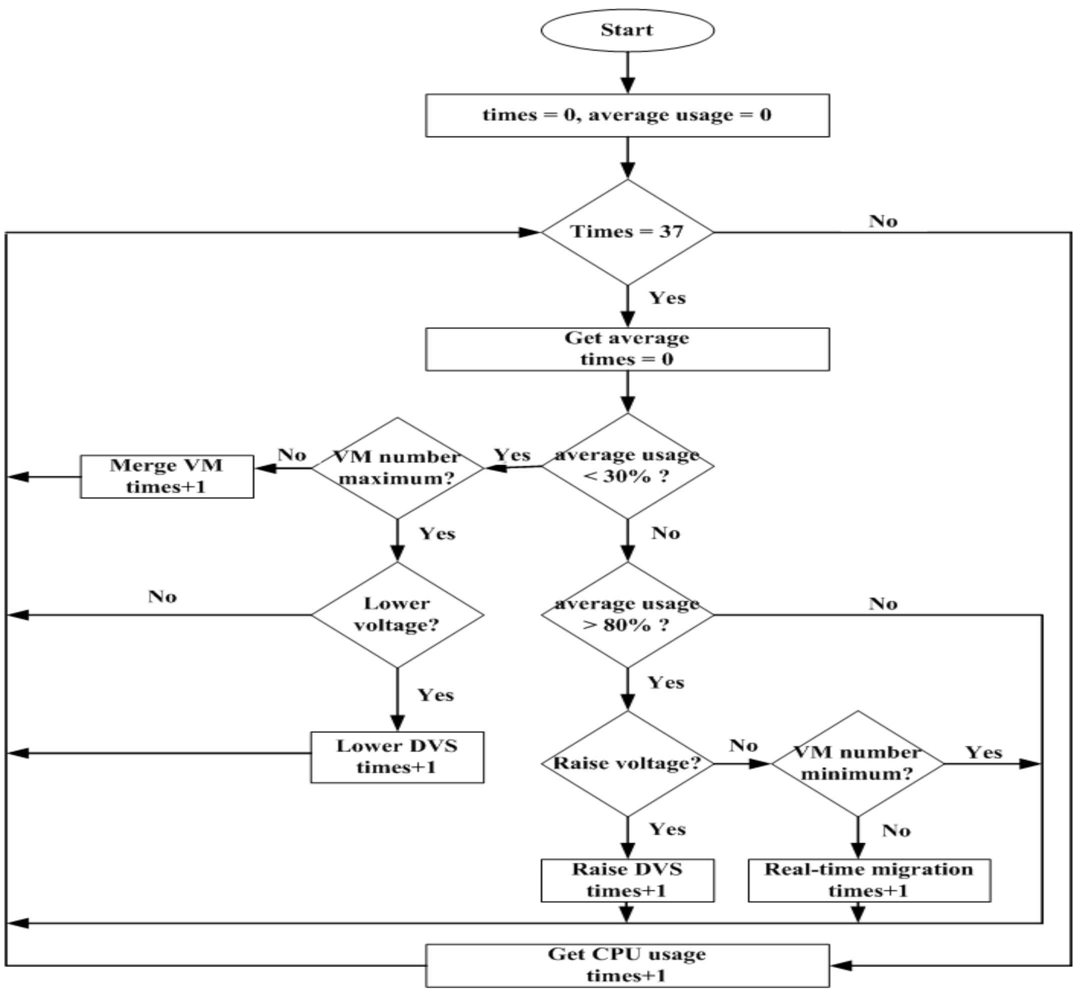

Figure 2. The System Flow Chart

3.3 The Resource Adjustment Mechanism

In the cloud computing environment, dynamic power consumption of the host can be calculated by Eq. 2.1. The average CPU utilization is recorded once every 5 seconds, and the voltage adjustment is determined every 15 seconds. The earliest recorded CPU utilization will be replaced by the latest every 5 seconds to update the average CPU utilization.

In the experiment, the voltage is adjusted to reduce the power consumption 25W at a time. In order to reduce the impact of continuously adjusting the voltage, voltage adjustment is performed if the CPU utilization is still too low or too high for a period. To lower the voltage will result in the degradation of the system performance. The relationship between the power consumption and system performance has been measured and listed in Table 1. The real-time migration can be performed for migrating the VM to other idle (or light workload) host, or merging VMs by migrating other VM to the light workload host in the experiment, however, a host is limited to support full utilization of 3 VMs at the same time.

Table 1. Power Consumption vs. CPU Performance

POWER (W)	CPU (MIPS)
250	247
225	228
200	207
175	182

4. Experimental Results and Analysis

In the experiment, the computer is using Intel® Core™2 Quad Q9500 CPU, running at 2.83GHz, 4GB DDR3 memory, and 500GB SATAII hard drive. Experimental platform is the CloudSim software with Windows 7 operating system. Both host and VM are single-core architecture. Mainly use the KVM, the compiler uses the IDE the Eclipse. A large scale cloud computing environment is simulated in the experiment. The number of tasks is 400, the resources required for each tasks are from 5 to 250 MIPS which are generated randomly. The voltage adjustment is determined every 15 seconds. Total run time is measured at the time all works completed. The experiment is performed 10 times for calculating the average value and total power consumptions.

Table 2. Average CPU Utilization vs. Energy Consumption

Average CPU Utilization (%)	DVFS mechanism applied (kwh)	Without applying DVFS mechanism (kwh)
25	0.58	0.63
50	0.66	0.69
75	0.76	0.75
100	0.84	0.81

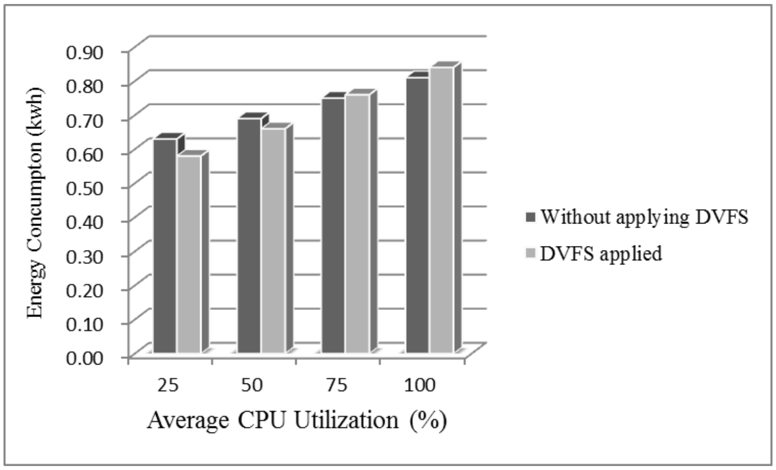

Figure 3. Energy Consumption Comparison

Table 2 and Figure 3 show energy consumption according to different average CPU utilization rates. It can be seen that when the CPU utilization rate is greater than 75%, energy consumption of the system with applying the proposed scheme will be slightly higher than that without applying the proposed scheme. The reason is that under the situation of heavy work load, unnecessary real-time migration will introduce extra energy consumption. When CPU utilization is less than 70%, the energy can be saved obviously by applying the proposed mechanism.

The relationship between the number of acting VMs and the migration and merging numbers is illustrated in Fig. 4. Increasing the number of acting VMs will result in increasing the number of migration and merging. In terms of energy efficiency, the more VMs can be placed on the same host, the greater the energy efficiency can be obtained. However, too many VMs on the same host will increase the number of migration and degrade the system performance.

Figure 5 shows the comparison of energy consumptions. The energy consumption without applying the proposed mechanism is presented as original, and DVFS presents the energy consumption measurement by applying the proposed mechanism. From the experimental result, we can see that about 24-33% of energy consumption can be saved by applying the proposed resources adjustment mechanism.

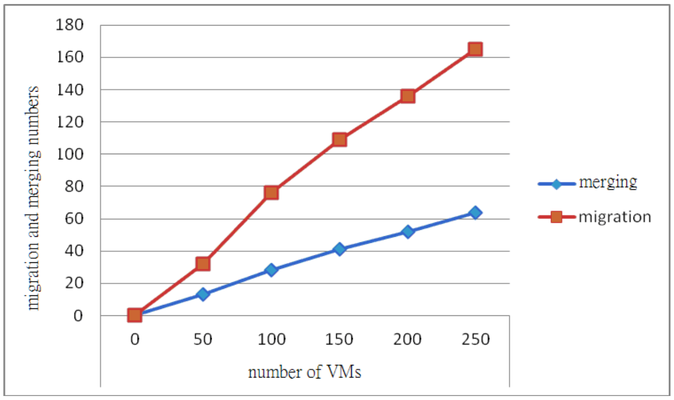

Figure 4. Number of Acting VMs vs. Migration and Merging Numbers

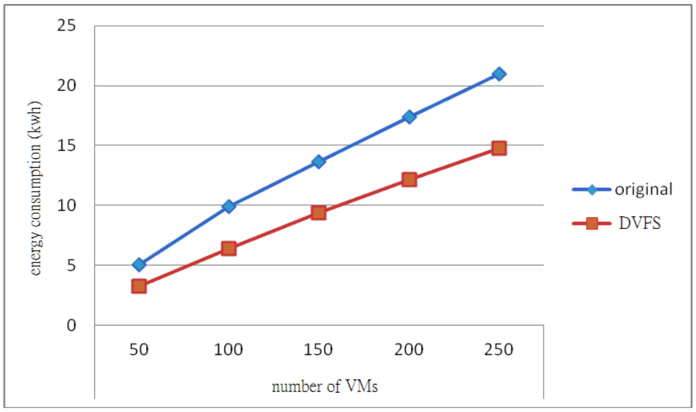

Figure 5. Energy Consumption Comparison

5. Conclusion

This paper proposed a dynamic resource management with energy-saving mechanism in the cloud computing environment to reduce the energy consumption. DVFS technique can be applied by monitoring the CPU utilization. When the workload is heavy, real-time migration can be provided for achieving more effective usage of resources under the user unaware situation. The experimental results show that the energy consumption can be saved significantly by applying the proposed dynamic resource management mechanism. Since a lot of migration will possibly lead to serious increasing of the extra energy consumption in the proposed architecture by applying the migration mechanism, in the future, multi-core architecture will be used to support multi-VMs. The number of acting VCPUs can be adjusted in accordance with the utilization of individual VCPU and the number of the CPU cores.

References

[1] R. D. Caytiles, S. Lee and B. J. Park, "Cloud Computing: The Next Computing Paradigm", International Journal of Multimedia and Ubiquitous Engineering (IJMUE), vol. 7, no. 2, SERSC, **(2012)**, pp. 297-302.
[2] Y. E. Gelogo and B. J. Park, "Next Generation Cloud Computing Issues and Solutions", International Journal of Control and Automation (IJCA), vol. 5, no. 1, SERSC, **(2012)**, pp. 63-70.
[3] H. Kim, H. Lee, W. Kim and Y. Kim, "A Trust Evaluation Model for QoS Guarantee in Cloud Systems", International Journal of Grid and Distributed Computing (IJGDC), vol. 3, no. 1, SERSC, **(2010)**, pp. 1-10.
[4] B. K. Kim, J. Kim and Y. W. Ko, "Efficient Virtual Machine Scheduling Exploiting VCPU Characteristics", International Journal of Security and Its Applications (IJSIA), vol. 6, no. 2, SERSC, **(2012)**, pp. 415-420.
[5] Y. F. Li, W. Li and C. F. Jiang, "A Survey of Virtual Machine System: Current Technology and Future Trends", Third International Symposium on Electronic Commerce and Security, **(2010)**, pp. 332-336.
[6] P. Barham, B. Dragovic, K. Fraser, S. Hand, T. Harris, R. N. A. Ho, I. Pratt and A. Warfield, "Xen and the Art of Virtualization", 19th ACM Symposium on Operating Systems Principles, **(2003)**, pp. 164-177.
[7] Y. Guo, X. Wang, W. Dong, G. Shi and Y. Li, "A cooperative model virtual-machine monitor based on multi-core platform", International Conference on Future Computer and Communication (ICFCC), vol. 1, **(2010)**, pp. 802-807.
[8] J. Pouwelse, K. Langendoen and H. Sips, "Dynamic Voltage on a Low-Power Microprocessor System", Proceedings of the International Conference on Mobile Computing and Networking, **(2001)**, pp. 251-259.
[9] N. Rodrigo, Calheiros, R. Ranjan, A. Beloglazov, C. A. F. D. Rose and R. Buyya, "CloudSim: A Toolkit for Modeling and Simulation of Cloud Computing Environments and Evaluation of Resource Provisioning Algorithms", ISSN:0038-0644, Wiley Press, New York, USA **(2011)**, pp. 23-50.
[10] Z. Liu, W. Qu, W. Liu and K. Li, "Xen Live Migration with Slowdown Scheduling Algorithm", International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT), **(2010)**, pp. 215-221.

[11] F. Ma, F. Liu and Z. Liu, "Live Virtual Machine Migration based on Improved Pre-copy Approach", IEEE International Conference on Software Engineering and Service Sciences (ICSESS), **(2010)**, pp. 230-233.
[12] G. Praveen, "Analysis of Performance in the Virtual Machines Environment", International Journal of Advanced Science and Technology (IJAST), vol. 32, SERSC, **(2011)**, pp. 53-64.

Authors

Liang-Teh Lee received the B.S. degree in Electrical Engineering from Tatung University in 1971 and his M.S. and Ph.D. degrees in Computer Engineering from University of Southern California in 1978 and 1989, respectively. He was the director of the Computer Center and the chairman of the Department of Computer Science and Engineering of Tatung University. Currently, he is the Professor of the Department of Computer Science and Engineering, Tatung University. His research interests include computer architectures, parallel and distributed computing, embedded systems, high speed switching architectures, and wireless networks.

Kang-Yuan Liu received the B.S. degree in Chemical Engineering from National Taiwan University of Science and Technology in 2002, and his M.S. degree in Computer Science and Engineering from Tatung University in 2004. Currently, he is the Ph.D. candidate in Tatung University. His research interests include video compression, real-time operating system scheduling, and embedded system programming.

Chia-Ying Tseng received a bachelor degree in industrial education from Taiwan Normal University in 1979 and a master degree in computer engineering from the Electrical and Computer Engineering Department of Arizona State University, USA, in 1985. In 1990, he was selected by Ministry of Education to enter the PhD program in the Computer Science and Engineering Department of Arizona State University for one year research. Currently, he is an Assistant Professor in the Computer Science and Engineering Department of Tatung University, Taiwan. His research interests include Embedded System Design, Embedded Real Time Operating System, Multiprocessor System-on-Chip (MPSoC), Cloud Computing and Virtualization.