Designing Parametric Constraint Based Power Aware Scheduling System in a Virtualized Cloud Environment

Harsimran Singh[1] and Harsimran Singh[2]

[1] Sri Guru Granth Sahib World University, Fatehgarh Sahib

Abstract

The increasing rate of the demand for computational resources has led to the production of largescale data centers. They consume huge amounts of electrical power resulting in high operational costs and carbon dioxide emissions. Power-related costs have become one of the major economic factors in IT data-centers, and companies and the research community are currently working on new efficient power aware resource management strategies, also known as ?Green IT?. Here we propose a framework for autonomic scheduling of tasks based upon some parametric constraints. In this paper we propose an analysis of the critical factors affecting the energy consumption of cloud servers in cloud computing and consideration to make performance very fast by using Sigar API to solve speed problems. In PCBPAS we impose some parametric constraints during task allocation to the server that can be adjusted dynamically to balance the server?s workloads in an efficient way so that CPU consumption can be improved and energy saving be achieved.

---

1 Introduction

he Cloud computing model leverages virtualization of computing resources allowing customers to provision resources on-demand on a pay-as-yougo basis [1]. Clouds are typified by utility access, largescale elasticity and rapid provision of resources, all of which are made possible through the use of virtualization on large IT clusters. In cloud computing CPU utilization & energy efficiency have become challenges to make it capable technology. Based on the trends from American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) **??**2], it has been estimated that by 2014 infrastructure and energy costs would contribute about 75%, whereas IT would contribute just 25% to the overall cost of operating a data center [3]. The reason for this extremely high energy consumption is not just the quantity of computing resources and the power inefficiency of hardware, but rather lies in the inefficient usage of these resources. Data collected from more than 5000 production servers over a six-month period have shown that although servers usually are not idle, the utilization rarely approaches 100%. Most of the time servers operate at 10-50% of their full capacity, leading to extra expenses on over-provisioning, and thus extra Total Cost of Acquisition (TCA). Moreover, managing and maintaining over-provisioned resources results in the increased Total Cost of Ownership (TCO).

The main aim of Energy-Aware Computing is to promote awareness of CPU & power consumption in both soft and hard systems. Clouds are built upon virtualization technology. The reduction in energy consumption can be achieved by switching idle nodes to low-power modes (i.e., sleep, hibernation), thus eliminating the idle power consumption as shown in Fig. 1. Efficient resource scheduling needs attention in a virtualized environment. This paper utilizes the predefined workload model in the virtual environment to propose the workload aware parametric constraints based power aware-scheduling (PCBPAS) technology, which is an extended form of automatic power saving technology called Power Aware Adaptive Scheduling (PAAS). The rest of this paper

is organized as followings. Section II discusses the related studies about virtualization migration policies, virtual machine migration and management. Section III proposes and explains PCBPAS methodology and its design issues. Section IV explains the simulation environment and procedures, also conducts the performance appraisal and analysis based on the simulation results. Finally, we make the conclusion and indicate the future research directions in Section V.

2 II.

3 Related Work a) Virtualization

Virtualization is a way to abstract the hardware and system resources from an operating system. It is a technology that allows running two or more operating systems side-by-side on just one PC or embedded controller [4]. It helps to better utilize available processing hardware to build more efficient systems. Most applications have only a finite amount of parallel tasks that can be executed at a given time, leaving many processor cores idle. Virtualization can solve this challenge by allocating group of processor cores to individual operating systems running in parallel. Thus benefits of virtualization can be enumerated as:i. Save hardware cost and footprint. ii. Take advantage of Operating System services. iii. Make use of Multi-core Processors.

Energy saving is a hot and one of core research areas in Cloud computing. On the one hand, people focus on reducing energy consumption of the data center infrastructures. For example, Tao Li's team provided an innovative idea that use renewable energy such as solar power in data center [5][6].

Another hand is using the virtual technology, such as energy consumption perception of virtual machine scheduling, and virtual machine transplantation [7][8].

And some scholars adopt optimized genetic algorithm to improve the performance of the resources schedule on physical machines [9].There is also one research on the relationship between the energy consumption and performance of data centers and the resource utilization, they see the resource allocation like the knapsack problem, which owns some special attributes such as disk utilization, CPU utilization and so on.

Song et al. [10] have proposed resource allocation to applications according to their priorities in multi application virtualized cluster. The approach requires machine-learning to obtain utility functions for the applications and defined application priorities. Unlike our work, it does not apply migration of VMs to optimize allocation continuously (the allocation is static).

Hsu-Yang Lin, [11] have proposed the Workload-Based APS scheme that automatically conducts the power distribution optimization in accordance with the predefined workload model. This technology mainly focused to migrate virtual machine from unloaded server to certain servers through the live migration and place the unloaded servers into the standby mode for reducing the server's power consumption. When the detector finds that workload is increasing from upper threshold, it will call wake-up function which will reactivate the standby server. The entire execution process does not require any interruption in the service and any manual operation. However, only two parameters lower and upper threshold are used in this approach. Moreover, these two parameters are static, which further reduces efficiency of the complete system, as it involves unnecessary VM migrations. Also, the migration cost of virtual machine is not considered in this approach.

4 III.

5 Problem Definition a) Problem Statement

The problem is to establish the technology required for achieving automatic power saving to balance the server's workload with some highly efficient and effective parametric constraints, so that it can lower migration costs and achieve power saving and finally reduce carbon emission. The aim of this research work is to impose some efficient and effective parametric constraints, during the task allocation that can be adjusted dynamically to balance the server's workloads in an efficient way so that CPU consumption can be improved and to achieve power, energy saving.

6 b) System Architecture

The major components in the purposed techniques are (Figure 2): i. Main Server Receives user request, process and classifies into smaller task units. These task units can be scheduled directly onto the JSP server but before that it needs to get assigned with random priorities. Priority can either be based on system state or the task characteristics. Once each task gets its unique priority these task units can be sent to the JSP server component to be scheduled.

7 ii. JSP Server

This server act as a scheduler. Scheduler pushes up the task units into appropriate schedule units based on the idleness and the saturation of each and every schedule unit. Scheduler units execute the task units based on the algorithm. The task unit with the lower deadline will be scheduled first to optimize the result. In this paper we have designed, workload based PCBPAS algorithm, which automatically conducts power distribution optimization in accordance with the predefined workload model. In general the requests of client are scheduled

to certain server i.e in idle condition or have low workload. But our purposed technology is mainly focused to schedule the requests of client to a server having normal or high workload based on the type of request instead of idle server, so that power of idle server can be saved. For the predictable system positioned in a certain low utilization rate of time interval, the scheduling can automatically execute task allocation and standby action of the server. Further, we also impose some parametric constraints during the task allocation, to improve the CPU utilization. To improve the CPU utilization task migration cost, we impose 3 parametric constraints given below: 1. No. of processes on server. 2. CPU idle rate. 3. Free memory space.

Every task given by the client will be scheduled to one of the two virtual servers according to these 3 parameters. The scheduling of the task based upon these parameters will be performed by JSP server. When the utilization rate did not reach the dynamic threshold value, the administrator would be automatically notified to adopt the corresponding power saving measure(s). Hence, the operations described above are illustrated in Figure 3.

8 d) Algorithm Design

From flow diagram it can be seen that, when client send any request to server, it will be received by Main server. This server act as classifier & it will classify the requests of different clients into smaller task units. If multiple clients sends requests to the server then it will assign priorities to each tasks based upon the type of task (longer or small). It will assign high priority to the smaller task low priority to the longer task. After classifying the task, it will send the task to the JSP server which acts as a scheduler & web server. This server will compute value of three parameters of two virtual servers A, B & compare it with each other. JSP server will always focus to assign the high priority task (smaller task) to the server with heavy workload instead of the server with small workload. represents the difference of no. of processes on both servers. After comparison, suppose if the free memory space of server A is lesser than server B it means the workload of server A is larger than B. Then it will check for idle rate of CPU of both servers. The server with lesser memory space will have lesser idle rate. So after checking idle rate of both servers it will check for the difference of no. of processes. If the difference of no. of processes is more than particular threshold then task of client will be allocated to the server having large no. of processes, lower CPU idle rate & lesser memory space that is server A. In this case server with lesser no. of processes, containing more memory space & high CPU idle rate behave as idle server that is server B. We can save more power of this server with better utilization of other server.

In accordance with the operational flow as depicted in Figure 2, the pseudo-code of the Workload-Based PCBPAS algorithm is designed as following: Algorithm Worklaod-Based PCBPAS BEGIN Set M FA , M FB as integer Set Idle_rate A , Idle_rate B as integer Set Processes A , Processes B as interger Set difference, threshold T as integer 1. When free memory space of A is less than memory space of B. 2. When idle rate of A is below the idle rate of B. 3. When the difference of no. of processes is above the particular threshold. Then JSP server will allocate the task to host A Host B will be placed in standby mode Otherwise JSP server will allocate the task to host B. Host A will be placed in standby mode. The platform actually constructed in this paper was based on the Netbeans which is an integrated development environment (IDE) for developing primarily with Java, but also with other languages, in particular PHP, C/C++, and HTML5. The NetBeans IDE is written in Java and can run on Windows, OS X, Linux, Solaris and other platforms supporting a compatible JVM. In our simulation environment, there is a cloud of five servers installed with window 7, out of which 2 are virtual servers, one is web server & one is set as common database, as illustrated in Figure 4. schedule high priority task to the server with smaller workload. In our technique scheduler always focuses to allocate the high priority task to the server that is already running with greater workload instead of that server that is running with very small workload. In our technique, the server with smaller workload acts as idle server. Using PCBPAS, the idle rate of idle server is always greater than idle rate when using PAAS technique. In this way, our technique has saved more power with better CPU utilization. The experimental results are shown below:- Now we can conclude from the Fig. 5 that the average idle rate using PCBPAS is larger than the idle rate using previous technique PAAS. In case of PCBPAS idle server remains 35% more in idle conditions as compared to the PAAS. The overall benefits associated with the additional reduced air-conditioning power consumption, which is attributed to the reduction in the heat ventilation when the host machine is in the standby mode, are not included.

9 Method

10 Global

V.

11 Conclusion and Future Scope

This paper gives the introduction of Cloud computing and background of various resource allocation and power management techniques. Our proposed workload-based PCBPAS can be able to achieve a workload balancing for the server and also provide an environment with better performance and improved utilization & power cost for the network or cloud platform. By analyzing the experimental results, the application of the workload-based PCBPAS achieve an approximately 35 % reduction in power consumption. Utilizing the workload-based PCBPAS can lower operating costs and achieve power saving with better CPU utilization.

158   Future opportunities could explore some more parameters for deciding scheduling requirement. Also a
159   technique that is both power-aware and thermal aware can be utilized to maximize energy savings both from
160   physical servers and the cooling systems used. Another direction for future research can be that the system is
161   made self learner or can be adjustable dynamically to improve the CPU usage & time cost of running server to
162   achieve power & energy saving.

163   # 12   Global Journal of Computer Science and Technology
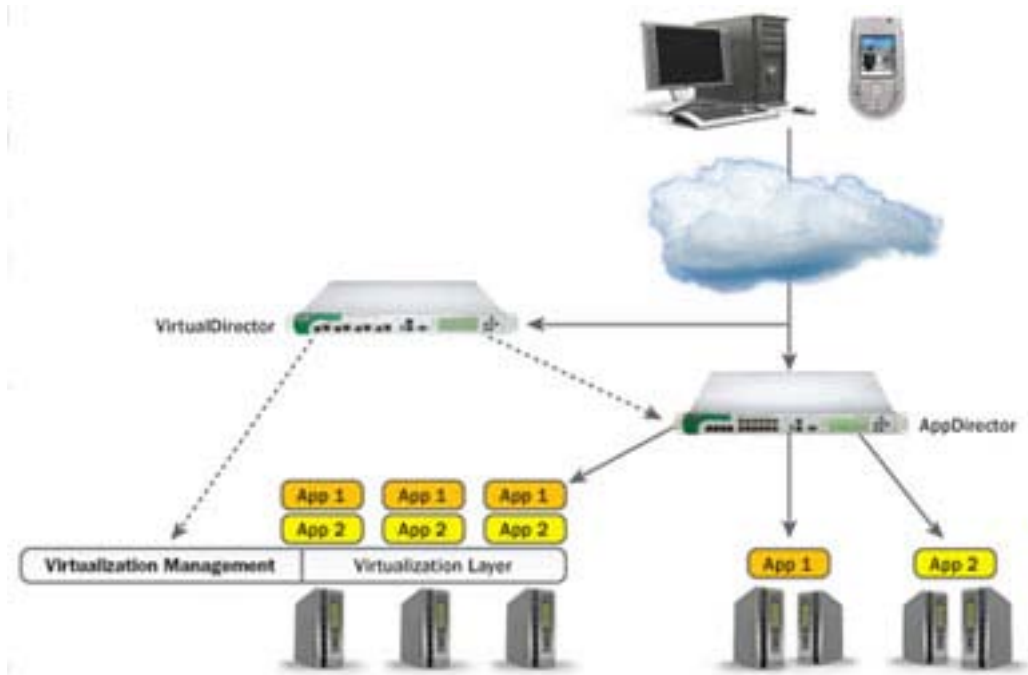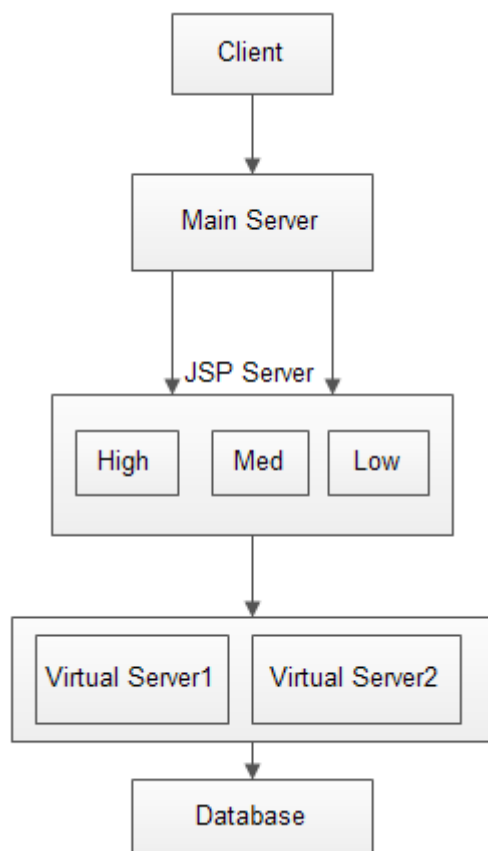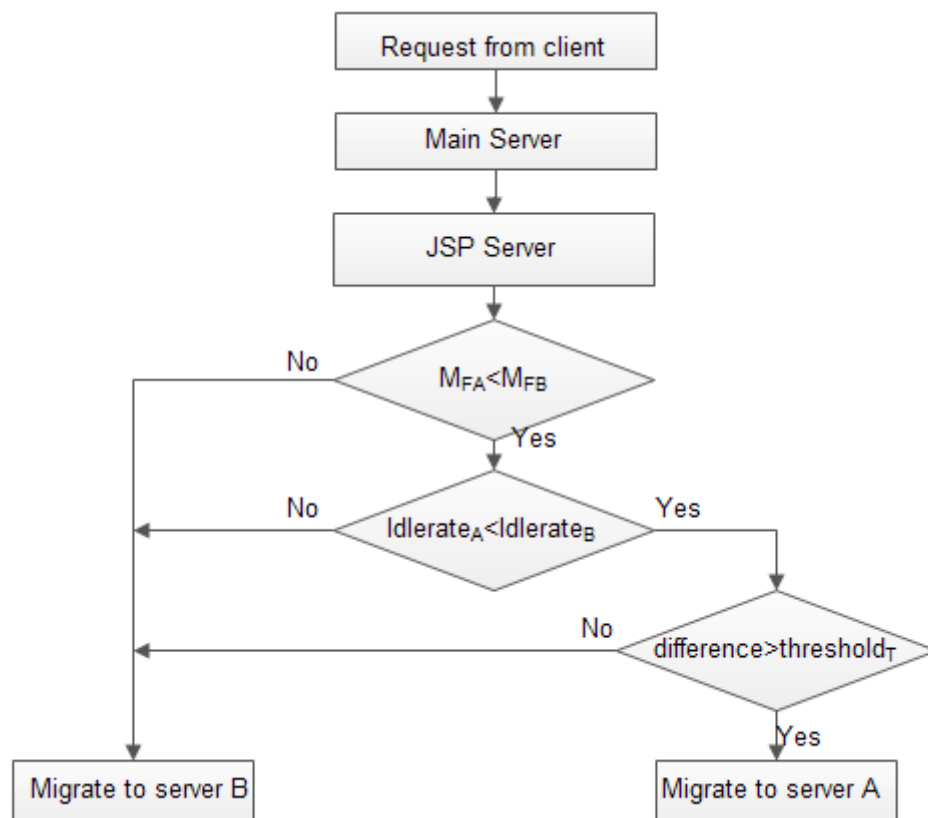

Figure 1: Figure 1 :

164   1

Figure 2: B

2

Figure 3: Figure 2 :

3

Figure 4: Figure 3 :

4
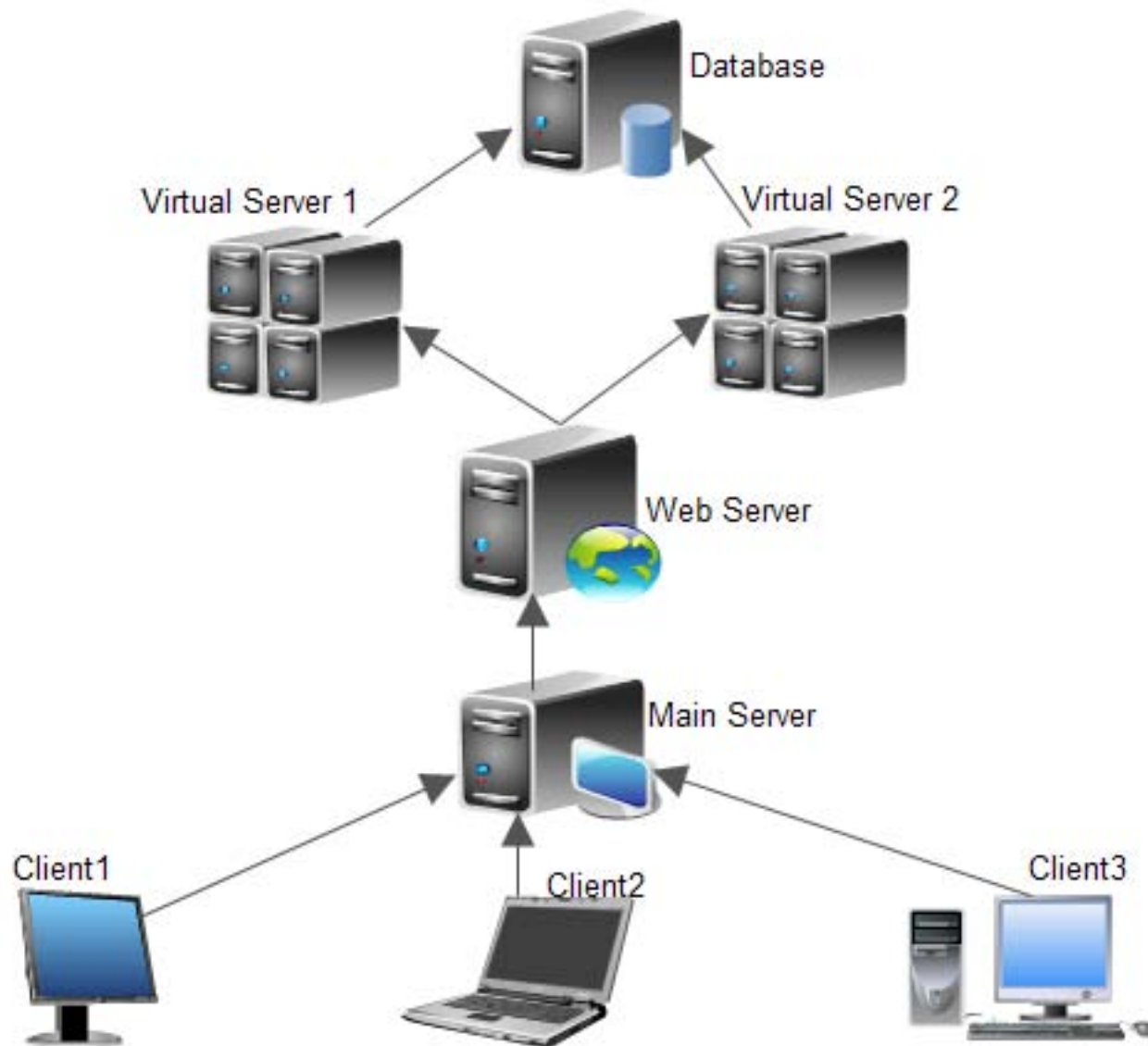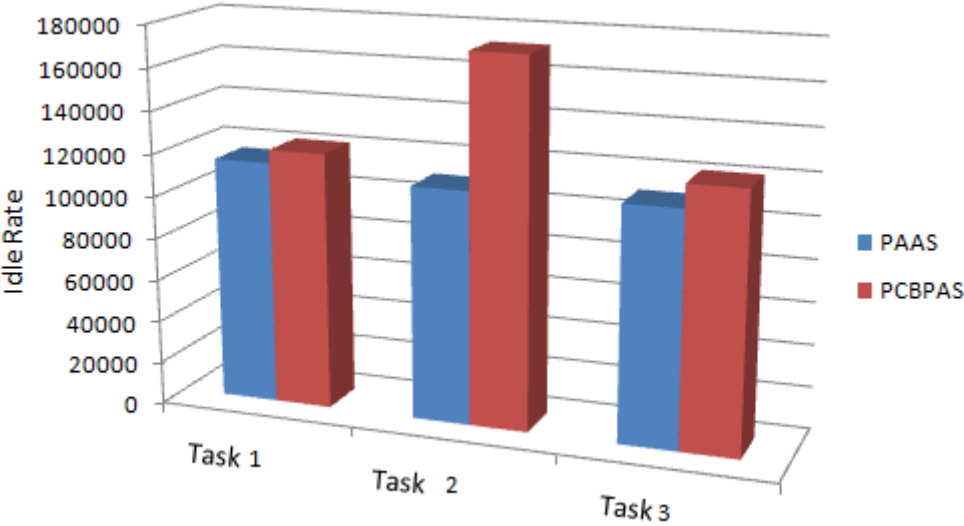
Figure 5: Figure 4 :

5

Figure 6: Figure 5 :

1

Task	Free Memory (Bytes)	Time Spent(Ns)	Idle Rate
1	199360976	0.000219	115384
2	201727232	0.000234	121644
3	203488701	0.000234	110087
4	200558175	0.000281	172852
5	202460108	0.000312	110451
6	204736101	0.000281	121562

Figure 7: Table 1 :

165 [Hsu and Feng] *A Power-Aware Run-Time System for High-Performance*, Chung-Hsing Hsu , Wu-Chun Feng .

166 [Zhong et al. ()] 'An Approach to Optimized Resource Scheduling Algorithm for Open-Source Cloud Systems'.
167   Hai Zhong , Kun Tao , Xuejie Zhang . *China Grid Conference (ChinaGrid'10)*, 2010. p. .

168 [Li et al. (2011)] 'Characterizing and Analyzing Renewable Energy Driven Data Centers'. Chao Li , Amer Qouneh
169   , Tao Li . *International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS*,
170   June 2011.

171 [Buyya et al. ()] 'Cloud computing and emerging IT platforms: Vision, hype, and reality for delivering computing
172   as the 5th utility'. R Buyya , C S Yeo , S Venugopal , J Broberg , I Brandic . *Future Generation Computer*
173   *Systems* 2009. 2005. 25 (6) p. . (B 2. ASHRAE Technical Committee 99)

174 [Computing Proceedings of the ACM/IEEE SC 2005 Conference ()] 'Computing'.   *Proceedings   of   the*
175   *ACM/IEEE SC 2005 Conference*, (the ACM/IEEE SC 2005 Conference) 2005. p. .

176 [Yang et al. ()] 'Designing Automatic Power Saving on Virtualization Environment'. Shin-Jer Yang , Lee-Chung
177   Chen , Hsi-Hui Tseng , Hui-Kuang Chung , Hsu-Yang Lin . *Proceedings of the International Conference on*
178   *Cloud and Cooperative Computing*, (the International Conference on Cloud and Cooperative Computing)
179   2010. p. .

180 [Younge et al. ()] 'Efficient resource management for Cloud computing environments'. A J Younge , G Laszewski
181   , Wang . *Green Computing Conference*, 2010. p. .

182 [Belady] *In the data center, power and cooling costs more than the it equipment it supports2007*, C Belady .
183   http://www.electronicscooling.com/articles/2007/feb/a3/

184 [Song et al. ()] 'Multi-Tiered On-Demand resource scheduling for VM-Based data center'. Y Song , H Wang , Y
185   Li , B Feng , Y Sun . *Proceedings of the 2009 9th IEEE/ACM International Symposium on Cluster Computing*
186   *and the Grid*, (the 2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid) 2009.
187   00 p. .

188 [Li et al. ()] 'SolarCore: Solar energy driven multi-core architecture power management'. Chao Li , Wangyuan
189   Zhang , Chang-Burm Cho , Tao Li . *High Performance Computer Architecture (HPCA)*, 2011. p. .

190 [Aasys (2008)] *Virtualization Basics*, Aasys . September 2008. 6.