ISSN 2278-3091
Volume 9, No.1.4, 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse0591.42020.pdf
https://doi.org/10.30534/ijatcse/2020/0591.42020

Energy Consumption Model for Virtual Machines in Cloud Data Centre

Mohd Zain Hassan[1], Hazilah Mad Kaidi[2], RudzidatulAkmam Dziyauddin[3], Shamsul Sarip[4], Harnani Hassan[5]

[1]Global IT Infrastructure and Operations, DXC Technology, HP Global Centre,Persiaran RimbaPermai, 63000 Cyberjaya, Selangor, Malaysia, mzainh@gmail.com
[2] Razak Faculty of Technology and Informatics, UniversitiTeknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia, hazilah.kl@utm.my
[3] Razak Faculty of Technology and Informatics, UniversitiTeknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia, rudzidatul.kl@utm.my
[4] Razak Faculty of Technology and Informatics, UniversitiTeknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia, shamsuls.kl@utm.my
[5]Faculty of Electrical Engineering, UniversitiTeknologi MARA, 40450 Shah Alam Selangor, Malaysia, harnani@uitm.edu.my

ABSTRACT

Cloud computing is a growing demand that allows a business to drive their goals using good Information Technology (IT). The tremendous increased in computing resources usage introduce several challenges to IT service providers. One of the key issues is the high energy consumption of the data centres. Virtualization has been used as a technique to reduce hardware footprint and thus realizing power efficiency in the data centre. However, improper placement technique has caused many virtual machines being provisioned inefficiently. Consequently, the physical servers installed as hosts are utilizing high power in data centre while the provisioned virtual machines are underutilized. This problem defeats the very reason why cloud computing was first introduced. This study focuses on identifying a relationship between virtual Central Processing Unit (CPU) of cloud servers, physical CPU, and total power consumption of its hypervisor. We developed a model to estimate power requirements for virtual machines (VM) in a cloud data centre and introduce an algorithm to place the virtual machines within the virtual data centre automatically. The model is applied to steady-state cloud data centres or in systems engineering design phase. The model can also be extended in determining a cloud server's chargeback model based on an individual could server proportional power consumption.

Key words: Cloud Computing; Data Centre; Green Computing; Power consumption; Virtual Machine Placement.

1. INTRODUCTION

Nowadays, the business strategy of companies is embedding information technology strategy in their business models [1] and thus making IT as one of the predominant factors in enabling business opportunities for an organization. The demand for computing resources has been increased tremendously over the last few years due to this scenario [2]. The trend has directly resulting growth and increase the need for businesses to subscribe to IT service providers [3]. The increased demand has challenged IT service providers to establish large-scale data centres containing many different types of computing equipment to cater to the needs [4].
Computing equipment, especially servers, are known for emitting carbon dioxide [2] from its operation and requires a cooling mechanism to lower down the temperature during operation [5]. Therefore, when a large scale of servers in data centres are confined, it will consume a high amount of electrical energy usage for its operation [3]. The recent advances in IT have introduced many techniques in managing efficiency of servers in data centres. Some of the techniques include physical server consolidation, physical server virtualization [6], natural air cooling system, and software-defined computing resource management. Virtualization concept introduces energy efficiency however, there isa lacking standard in measuring virtual machine and host power efficiency in the data centre[7]. It introduces challenges to understand the deep and complex interactions between the performances of virtual systems with power management and how efficiency can be managed [8].

Many IT service providers tend to over-provision virtual machines in the cloud data centre, assuming that energy efficiency is achieved through virtualization [7]. However, this introduced a new problem when high energy is consumed by physical hosts that are not fully utilized [9]. In addition, virtual machines placement that ais not balanced by its energy usage also contributes to the problem [10]. The main contributor to this problem is due to the lack of standard in estimating virtual machine energy consumption [7].

A mathematical model is developed to estimate the power consumption of hypervisor-based on virtual servers' utilization from this research. The model is then used together with the classification data from manufacturing standard to propose an algorithm that will help the IT Service Providers in achieving energy efficiency in the cloud or virtualization systems based on the virtual server's utilization and power estimation model.

2.METHODOLOGY

The data collection in this research illustrates in Figure 1. Physical layers involve data from every host in the cloud data centre farms in this research. There are 16 hosts configured for each of the farms. Each of the host's physical hardware specification is explained in Table 1. The power consumption and CPU utilization data were collected for each machine at 20 seconds interval.It is important to take note that only server with Energy Star rating should be considered to ensure power-efficient hardware are being used, as shown in the table.

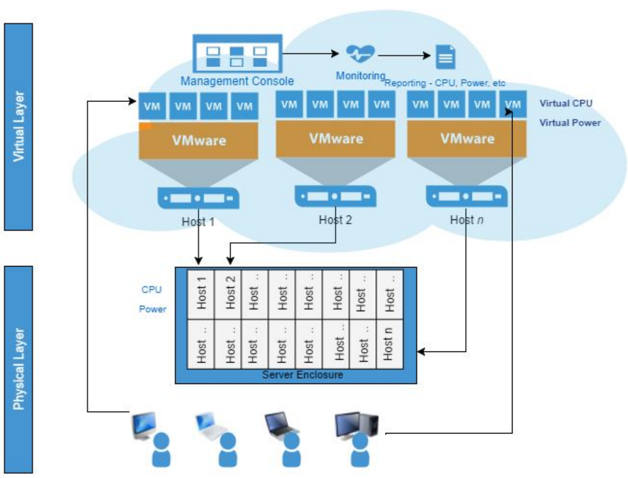


Figure1: Conceptual Framework of the Study

Table2: Specification of the physical hardware

Manufacturer	Hewlett-Packard Company
Model	ProLiant BL465c Gen8
Energy Rating	Energy Star
Power Profile	No power profile set
Minimum Power	129 W [11]
Maximum Power	500 W [11]
Processor(s)	24 x AMD Opteron(TM) Processor 6238. Bus speed 2400 MHz
Memory	512 GB
Operating Systems	VMware ESXi 5.5

The minimum and maximum power requirement was provided by the manufacturer whitepaper based on the power input of 240V. In the virtual layer, the data is taken by getting the CPU utilization reading in every virtual machine hosted in 16 hosts. There is a total of 938 virtual machines involved during the data collection. The tools and software used for the collection will be further explained in the following section.

Data centre efficiency is measured using power usage effectiveness (PUE) indicator [12]. It is measured as a ratio of IT load against the total facility load. The goal in data centre efficiency is to reduce the PUE to the lowest possible value. The IT load is referring to the power usage of IT equipment comprises of servers, network, storage devices and management hardware [5]. The work focuses on the server element to reduce total power consumption. In the data centre, the total server power is the summation of non-hypervisor and hypervisor server power. Within the organization where this study is conducted, 90% of the workload is hosted by hypervisor servers. According to study [13], [14], power consumption and CPU utilization have an almost linear correlation between two machines. Although some servers are applying Dynamic Voltage Frequency Scaling (DVFS), the study [13] reveals that the method does not impact much in predicting power consumption, thus linear correlation can still be used effectively. In this study, a linear correlation of power model is proposed by referring to cloud server utilization. It is also understood that Operating Systems of the hypervisor server has its kernel and management processes that utilize a certain amount of available CPU. All this component can then be written as a hypervisor power, *HP* described in (1) where $HK_{cpu}$is a hypervisor kernel CPU usage.

$$HP = VMs\ virtual\ CPU + HKcpu \quad (1)$$

Several research questions have been developed to design data collection and analysis for this study. The data will be collected from one of a leading global IT Service Provider company based in the USA. Firstly, data of server utilization during idle and peak performance is gathered from server hardware manufacturer whitepaper to answer the question of host minimum and maximum power consumption. Secondly, another important question is to understand the relationship between host CPU utilization and its power. The third question is to understand the relationship between virtual server's CPU usage and its hypervisor CPU. Both questions are addressed by collecting performance data of 16 hypervisors and 937 virtual servers over a period of 5 days in every 20 seconds interval for one hour.

All the hypervisors in this study are using VMware Operating Systems. A custom script is written to collect the host CPU data in the mentioned period utilizing Microsoft PowerCLI software. Hypervisor host power information is collected by running a custom script using PowerShell software. Power supply sensor of the server is queried directly using manufacturer API tool called HPiLO cmdlet. This utility is installed as a component in Microsoft PowerShell software in a central location server located within the same data centre of servers involved. This is to prevent potential network congestion which may cause an error during the collection. An automated job is set so the data can be collected simultaneously in the mentioned period. After five days, the data is extracted and combined using Microsoft Excel. The combined data is then analyzed using statistical software Minitab to get the required correlation and regression model.

3.RESULT AND DISCUSSION

3.1 Relationship of Virtual Machine CPU, Host CPU and Host Power Utilization

The relationship between host CPU utilization and power consumption is illustrated in Figure 2. By using Pearson analysis, both variables are tested, and the P-Value is determined as 0.01. Since the P-Value is low (less than 0.05), the study indicates that the correlation is statistically significant based on a 95% significant level. A linear correlation in Figure 2 implies a fitted equation as in (2).

Power utilization = 194.6 + 0.0003766 x CPU usage (2)

Figure 2 yields that the host power utilization can be predicted from the CPU usage. The results are compared against a power model introduced by Beloglazov, 2013 [9].

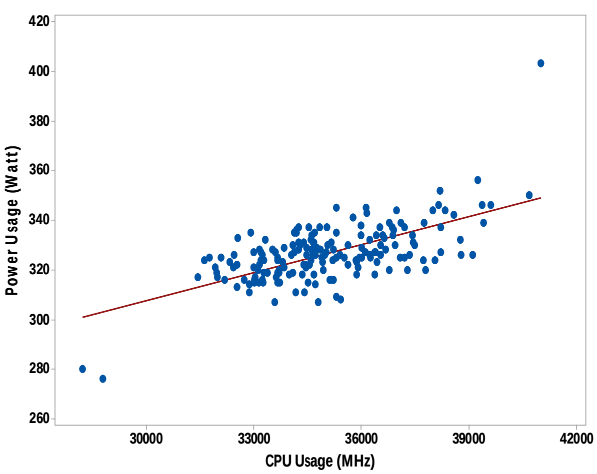
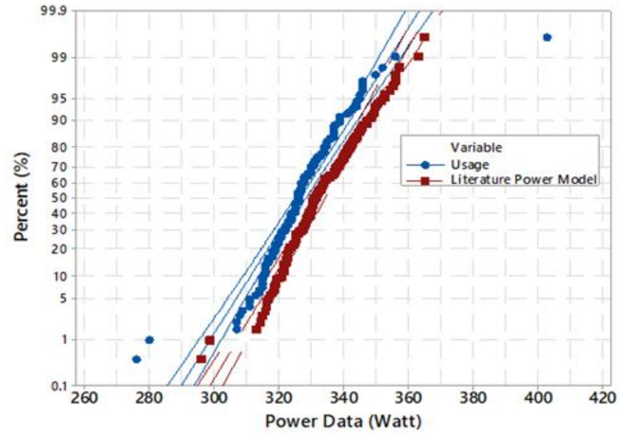


Figure2: Analysis of Relationship between Hypervisor CPU and Power Utilization

The collected data is fitted into the power model in the previous study. A comparison of the proposed power model and the previous power model [6] using a probability plot with a 99% confidence level is shown in Figure 3. The power model in previous study is labeled in red dotted which proportional with the proposed model. Figure 3 demonstrates a clear linear pattern for both power models, which validates the proposed power model in this study. In essence, the proposed power model can be used to describe the relationship between CPU usage and power utilisation for hypervisors.



In order to analyze the relationship between total virtual machines and host CPU usage, a linear regression analysis is performed. The results of the analysis are presented in Figure 4. It can be noted from the results that there is a significant positive relationship between totalVM and host CPU usage. A third variable, host kernel processes are also explored, and a regression model is developed with strong R-sq variation of 99.80%. The model is explained in (3).

$$hCPU = 1067 + 0.99267 * \sum_{i=1}^{n} vCPU + 1.0171 * kCPU \quad (3)$$

where *vCPU* is the total CPU usage of the virtual machines, and *kCPU* is the CPU utilization of the kernel processes in the host.

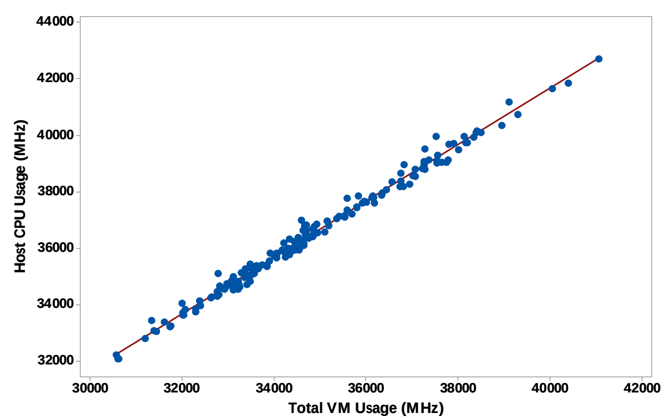


Figure4: Total VM Usage and Host CPU Usage Relationship

3.2. Identifying Over-utilized Host and Under-utilized Host

The over-utilized host is a condition of a server when its load has reached more than its upper limit [15]. Typically, the upper limit of the hypervisor can be set by any performance parameter of a server. In this research, power consumption is taken as the parameter to determine the condition of a hypervisor to optimize power efficiency. The maximum power limit of server is determined by the hardware manufacturer and can be simulated using simulation software or by running a workload that utilized 100% of CPU frequency. In this study, the upper limit is set to 90% from its actual peak power to give 10% allowance of a possible power surge that may cause performance degradation of the server.

Technical whitepaper by VMware, 2013 [16] and study by Beloglazov, 2013 [9] mentioned that servers are considered as inefficiency when the resources are less than 50% utilized in an average within a time. In such a case, the server is called under-utilized server [15]. In this study, CPU is the factor that determined power consumption; the goal in identifying under-utilized host is to find the power consumption and increase its utilization to more than 50% from maximum allowable CPU utilization. From Table 1, it is known that the hypervisor server is having a maximum of 57200 MHz of CPU frequency. In order to optimize the CPU at least 50% of the usage, the power model P_{low} is obtained as in Equation (4).

$$P_{low} = 194.6 + 0.003766 \times (0.5 \times 57200) = 302\ W \quad (4)$$

3.3Virtual Machine Placement

Two algorithms are proposed to tackle host under-utilized and over-utilized conditions. In Algorithm 1 shown in Table 2 below, the threshold value discussed in the previous chapter is referred to as P_{max}. Host power utilization is captured for a period of 10 minutes, and its mean value is compared against the power threshold value. When the mean power is found as higher than the threshold, the host status is set to true to indicate the host is over-utilized. Post that, the CPU utilization for each of the virtual machines hosted in the host are collected, and each machine is compared against the quartile three value of the total virtual machine's CPU. The virtual machines which utilize the top 25% of CPU resources in the hosts are then selected to be moved to other hosts.

Table1: Host Over-utilized Detection Algorithm

Algorithm 1: Host over-utilized detection algorithm

```
Input: utilization, threshold, n, status, vmsMap, sumVMcpu
Output: Whether the host is over-utilized
ifutilization is not empty then
    utilization← last n values of utilization
    meanUtilization← sum(utilization)/ len(utilization)
    ifmeanUtilization> threshold then
        status = TRUE
return status
if status is TRUE then
selectedVm[]← None
  for vm in vmsMapthen
    ifvmsMap[vm].cpuUtil> quartile(sumVMcpu, 3) then
        selectedVm←vm
  returnselectedVm
```

Algorithm 1.1: VM Placement for Over-utilized Host

```
Input: selectedVm[], hostsFarm[], curHost
Output: New host for the selected virtual machines
hostsFarm[] ←sortbyLowestPower()
forhost in hostsFarmdo
  if host.status is FALSE AND host is NOT curHostthen
    ifcpu< sum(selectedVM[].cpu) then
      eligible←estimatePower (host.cpu+sum(selectedVM.cpu),host.kernel)
      ifeligible is TRUE then
      for each vm in selectedVMdo
          vm.NewHost← host
       returnvm
        break for
    else
      for each vm in selectedVMdo
       newHostCPU←host.CPU + vm.CPU
eligible←estimatePower(newHostCPU,host.kernel)
          ifeligible is TRUE then
              vm.NewHost← true
          ifeligible is FALSE then
              Exit for and move to next host
selectedVm←vm
 returnselectedVm
```

Algorithm 1.1 is then called to perform target host selection. A new host is selected by checking its current power starting from the host with lowest power reading. Power prediction is performed using the power model from this research to determine if the new host would be able to host the VMs selected for migration. Once the new host is finalized, the VMs will be moved, and new data collection will be repeated in 10 minutes cycle.

Algorithm 2 is developed to detect the host with an under-utilized condition, as shown in Table 2. The P_{low} value obtained from the estimation calculated in section 3.2 is used as the threshold value in the algorithm. The algorithm is run for every 10 minutes for each host in the farm. It will detect host that is having mean utilization in the 10 minutes time of less than the threshold value. The logic applied in this algorithm ensures that hosts CPU are optimized to 50% and more of its capacity. Whenever the algorithm detects the under-utilized host, it will move out all the virtual machines from the hosts and execute virtual machines placement to all other hosts that are not in under-utilized or over-utilized condition. The virtual machines placement will ensure the target host is optimized and not taking too much of load until it becomes an over-utilized host. This placement is done automatically through the HA features of VMware vSphere. After all virtual machines have been moved out, the host is then put into a sleep, so power efficiency is applied.

Table-2 Host Under-utilized Detection Algorithm

Algorithm 2: Host under-utilized detection algorithm

```
Input: utilization, threshold, n
Output: Whether the host is under-utilized
ifutilization is not empty then
   utilization← last n values of utilization
   meanUtilization← sum(utilization)/ len(utilization)
   ifmeanUtilization< threshold then
      returntrue
return true
```

Algorithm 2.1: Host under-utilized detection algorithm

```
Input: selectedVm[], hostsFarm[], curHost
Output: New hosts for selected virtual machines
hostsFarm[] ←sortbyLowestPower()
forhost in hostsFarmdo
  if host.status is FALSE AND host is NOT curHostthen
     ifcpu< sum(selectedVM[].cpu) then
        eligible←estimatePower
(host.cpu+sum(selectedVM.cpu),host.kernel)
        ifeligible is TRUE then
        for each vm in selectedVMdo
           vm.NewHost← host
         returnvm
          break for
     else
        for each vm in selectedVMdo
newHostCPU←host.CPU + vm.CPU
        eligible←estimatePower(newHostCPU,host.kernel)
           ifeligible is TRUE then
              vm.NewHost← true
           ifeligible is FALSE then
              Exit for and move to next host
selectedVm←vm
 returnselectedVm
sleepHost (curHost)
```

The algorithm of virtual machine scheduling can be improved using Hyper Heuristic approach asin [17]and can be implemented and evaluated using Hybrid Meta-heuristic in [18].

4.CONCLUSION

This studywas conducted to establish the relationship of virtual machines CPU usage to host CPU and its power consumption. The regression analysis has revealed that there is a significant relationship between the virtual machine's CPU and host CPU usage. The relevance of this relationship was also supported by the regression analysis of host CPU usage to its power. Based on the results of the multi regression analysis, a simplified power model of the cloud data centre farm was developed to predict the power consumption of a host. For future work, the power consumption model will be conceptually tested, and the relationship of host CPU usage, host CPU and power consumption can be accessed by using machine learning.

ACKNOWLEDGEMENT

This work is supported by UniversitiTeknologi Malaysia under Research University Grant Q.K130000.3556.06G45 for the financial support provided throughout this research project.

REFERENCES

1. A. Bharadwaj, O. a. El Sawy, P. a. Pavlou, and N. Venkatraman.**Digital Business Strategy: Toward a Next Generation of Insights**, *MIS Q.*, vol. 37, no. 2, pp. 471–482, 2013.
2. J. Subirats and J. Guitart, **Assessing and forecasting energy efficiency on Cloud computing platforms**, *Futur. Gener. Comput. Syst.*, vol. 45, pp. 70–94, 2015.
3. E. Arianyan, H. Taheri, and S. Sharifian, **Novel energy and SLA efficient resource management heuristics for consolidation of virtual machines in cloud data centres,***Comput. Electr. Eng.*, vol. 47, pp. 222–240, 2015.
4. F. Kong and X. Liu, **A Survey on Green-Energy-Aware Power Management for Datacentres**, *ACM Comput. Surv.*, vol. 47, no. 2, pp. 1–38, 2014. https://doi.org/10.1145/2642708
5. A. H. Beitelmal and D. Fabris, **Servers and data centres energy performance metrics**, *Energy Build.*, vol. 80, pp. 562–569, 2014.
6. A. F. Monteiro, M. V. Azevedo, and A. Sztajnberg, **Virtualized Web server cluster self-configuration to optimize resource and power use**, *J. Syst. Softw.*, vol. 86, no. 11, pp. 2779–2796, 2013.
7. M. Uddin and A. A. Rahman, **Energy efficiency and low carbon enabler green IT framework for data centres considering green metrics**, *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 4078–4094, 2012.
8. K. W. Cameron, **Energy efficiency in the wild: Why datacentres fear power management**, *Computer (Long. Beach. Calif).*, vol. 47, no. 11, pp. 89–92, 2014.
9. A. Beloglazov, **Energy-Efficient Management of Virtual Machines in Data Centres for Cloud Computing**, PhD thesis, Department of Computing and Information Systems, The University of Melbourne, 2013.
10. X. Li, Z. Qian, S. Lu, and J. Wu, **Energy efficient virtual machine placement algorithm with balanced and improved resource utilization in a data centre**, *Math. Comput. Model.*, vol. 58, no. 5–6, pp. 1222–1235, 2013.
11. U. S. E. P. Agency, **ENERGY STAR Certified Enterprise Servers | EPA ENERGY STAR**, *Energy Star*, 2016. [Online]. Available: https://www.energystar.gov/productfinder/product/certified-enterprise-servers/. [Accessed: 29-Dec-2016].

12. V. Avelar, D. Azevedo, and A. French, **PUE^{TM}: A Comprehensive Examination of the Metric**, The Green Grid white paper, pp. 1–83, 2012.
13. H. Rong, H. Zhang, S. Xiao, C. Li, and C. Hu, **Optimizing energy consumption for data centres**, *Renew. Sustain. Energy Rev.*, vol. 58, pp. 674–691, 2016.
https://doi.org/10.1016/j.rser.2015.12.283
14. A. Beloglazov and R. Buyya, **OpenStack Neat: A framework for dynamic and energy-efficient consolidation of virtual machines in OpenStack clouds**, *Concurr. Comput.*, vol. 27, no. 5, pp. 1310–1333, 2015.
15. Z. Usmani and S. Singh, **A Survey of Virtual Machine Placement Techniques in a Cloud Data Centre**, *Procedia Comput. Sci.*, vol. 78, pp. 491–498, 2016.
16. Qasim Ali, **Host Power Management in VMware vSphere 5 Performance Study**, Technical white paper VMware CA, 2013.
17. Ahmad Alauddin Ariffin, Aws Fadhil Ibrahim, Sazlinah Hasan and Rohaya Latip.**An Efficient Virtual Machine Scheduling Algorithm to Minimize Makespan and Maximize Profit Using Hyper Heuristic Approach,** International Journal of Advanced Trends in Computer Science and Engineering**,**Vol 8, 206-216, 2019.
https://doi.org/10.30534/ijatcse/2019/3281.42019
18. Saif, Faten and Alwan, Ali. **Performance Evaluation of Task Scheduling using Hybrid Meta-heuristic in Heterogeneous Cloud Environment**, International Journal of Advanced Trends in Computer Science and Engineering. Vol 8, 3249-3257, 2019.
https://doi.org/10.30534/ijatcse/2019/93862019