Intelligent Agent-based Technique For Virtual Machine Resource Allocation For Energy-Efficient Cloud Data centres

OGECHUKWU M OKONOR, MO ADDA, ALEX GEGOV
School of Computing
University of Portsmouth
UNITED KINGDOM

Abstract: -In recent years, cloud computing technology has emerged as a promising solution to many medium and small scale business enterprise. It leverages cost, encourages collaboration, and it is so flexible to use. Despite its attractive features, it still has some challenging issues such as power consumption rate. Although significant efforts have been devoted to optimising power usage through scheduling and virtualisation, it though remains a challenging problem. This paper introduces a new model on how to reduce the power consumption during live virtualisation and schedule based on intelligent mobile agent technology. The simulation results show that the proposed algorithm minimises power consumption used in the data centre at each stage by more than 30% compared to the current state-of-the-art techniques.

Key-words: energy efficiency, virtualisation, intelligent agent, cloud computing, data centre

1 Introduction

Data centres (DC) are structured electronic components which are responsible for its data processing (servers), its data communications (network component), and its data storage. These computing and networking components aid the processing, storing, and transmitting digital information in the information technology world. There has been a significant advancement in the rate at which people embrace technology. The adoption rate has become increasingly high due to recent innovations in smarter electronic gadgets. This digital transformation directly affects the DC performance, and inversely affects the energy consumption rate through its activities. Although some of the DC equipments were not originally designed to be active but has little or no heat resistor embedded in it. According to [1], real-time video streaming, online gaming, as well as mobile device are all on the increase. They contribute about 60% of all data traffic in cloud DC activities with a prediction of 80% rise by 2020.

Furthermore, the service level agreement (SLA) binding the service providers and their clients DC runs 24/7 all year round with a typical power density of 538-2153w/m^2 [2]. The energy consumed by the DC is now so alarming.

In 2007, according to the report presented to the US congress shows that US data centre consumed about 61 billion kilowatt-hours in 2006, which is about 1.5% of the total U.S electricity consumption [3]. The Garter report published three years later showed a further 0.5% increase [4].

In 2016, some of the data centre grant vendors claimed to be in control of the skyrocketing level of DC energy consumption. According to the U.S data centre energy usage report by Berkeley national laboratory [5], it claimed that due to the increasing awareness in data centre infrastructure operation, significant improvement has so far been seen in the energy efficiency of data centre structure. This is attributed to the adoption of hyperscale data centres which encourages a high level of virtualisation and reduces the use of physical servers. However, researchers and other industry experts have some doubts on this due to other research findings on energy efficiency involving cloud data centres.

While the US report estimates a static growth on its energy efficiency, the world is still struggling on the amount of energy used by the data centre systems which shows there is some level of hidden truth yet to be told by some of the vendors. According to the energy watch manger publication in 2019 [6], it shows that 4.7% of annual electricity usage globally is from the data centre.

Therefore, systematic observation has shown that a more dynamic method with intelligent operation should be the headway to finding a lasting remedy to high power usage in the cloud data centre. Before this study, the option of choosing the best energy consumption model suited to the needs of the data centre’s high energy usage level has remained a

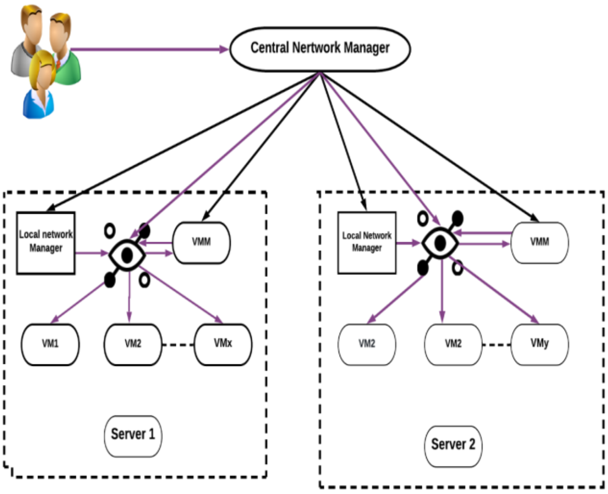


Figure 3: System Model

In this system model, the server has a multicore CPUs. Each multicore CPU has K cores, and each core has M MIPs as a single-core which make the CPU sum up its total capacity as K*M MIPS. Table 1 below gives a clear description of the server specification and configuration used in this research work.

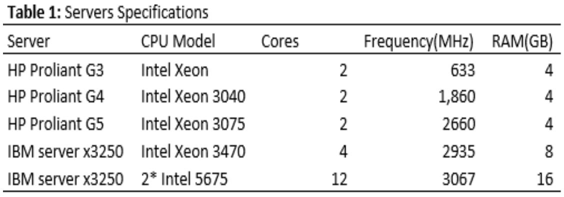


Table 1: Servers Specifications

Server	CPU Model	Cores	Frequency(MHz)	RAM(GB)
HP Proliant G3	Intel Xeon	2	633	4
HP Proliant G4	Intel Xeon 3040	2	1,860	4
HP Proliant G5	Intel Xeon 3075	2	2660	4
IBM server x3250	Intel Xeon 3470	4	2935	8
IBM server x3250	2* Intel 5675	12	3067	16

4.3 Experiment setup

In this research work, Cloudsim toolkit is used because it is an extensible open-source toolkit with built-in capability to compare power-aware algorithm with its performance metric. Cloudsim also models the real-time of the federated cloud environment and is now the most acceptable platform for conducting cloud base simulation. It features support virtualisation engine which enables the creation of adequate management of independent, multiple and virtualised services on the data centre host.

The experiment was set up for 10 days with data from CoMon project [18] with more than 600 heterogeneous physical machines, monitoring infrastructure from PlanetLab. The data contains the CPU utilisation with 5-minute interval recording from more than 1000 VMs. The workload traces is an infrastructure-as-a-service cloud model dataset, which is ideal for this piece of work. It also modelled Amazon elastic compute cloud (EC) and IBM specification. We used a random workload with random scheduling technique with some intelligent embedded message accessing all data centre. In each scenario, each VM runs a variable workload with agent monitoring and intelligent messaging technology to ensure efficiency. Table 2 shows the power consumption of the chosen server specification in Table 1 based on a 90% CPU utilization rate using 90% utilisation rate leverage automatic downtime due to overload.

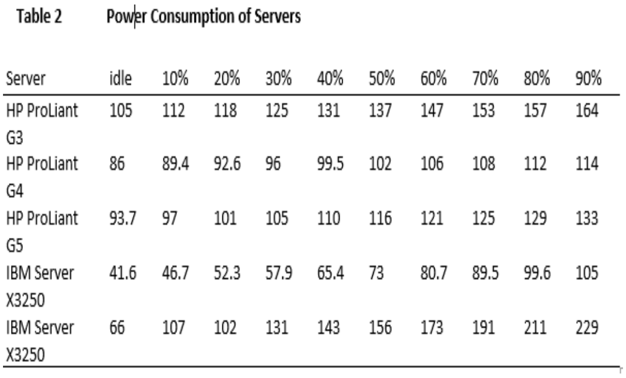


Table 2 Power Consumption of Servers

Server	idle	10%	20%	30%	40%	50%	60%	70%	80%	90%
HP ProLiant G3	105	112	118	125	131	137	147	153	157	164
HP ProLiant G4	86	89.4	92.6	96	99.5	102	106	108	112	114
HP ProLiant G5	93.7	97	101	105	110	116	121	125	129	133
IBM Server X3250	41.6	46.7	52.3	57.9	65.4	73	80.7	89.5	99.6	105
IBM Server X3250	66	107	102	131	143	156	173	191	211	229

For any performance in the system to be valid, it must consider the SLA violation (VSLA) metric. We consider for each approach the SLA violation that occurs during the process and the Performance degradation Management level (PDML) of each transaction caused by the system. We then model as follows -

1: SLA violation per active server(SLATASv): where CPU utilization reached 100%.

$$SLATASv = \frac{1}{N_{sv}} \sum_{i=1}^{N_{sv}} \frac{T_{si}}{T_{\alpha i}} \tag{16}$$

where

N_{sv}: number of servers

T_{si}: total time it took the server to reach 100% utilisation

$T_{\alpha i}$: active server per time

2: Performance degradation management level is observing the live migration of VM from one server to the other without obstruction. Degradation can be affected by many application behavioural factors. The estimated performance degradation management level on the CPU at 10% utilization rate is as follows:

$$PDML = \frac{1}{N_{Vms}} \sum_{j=1}^{N_{vm}} \frac{C_{dj}}{C_{rj}} \tag{17}$$

where

N_{Vms}: $is\ the\ number\ of\ vms$

C_{dj}: estimated as 10% cpu utilization of vmj in all migration

C_{rj}: $total\ cpu\ requested\ by\ vmj$

Hence, the equation for formulating VSLA is then calculated by combining the two metrics in Eqn 16 and Eqn 17.

$$VSLA = SLATASv * PDML \qquad (18)$$

Furthermore, for clarity purpose, we defined the characteristic, Mips and workload data in Table 3.

Table 3 **Virtual Machine (EC2)**

VM Type	**CPU(MIPS)**	**RAM(GB)**
High-Memory Extra Large	3,000	6,000
High-CPU Medium	2,500	850
Extra Large Instance	2,000	3750
Small Instance	1,000	1,700
Micro Instance	500	633

Table 4 represents the conducted parameters of workload calculated over 10 days of the experiment.

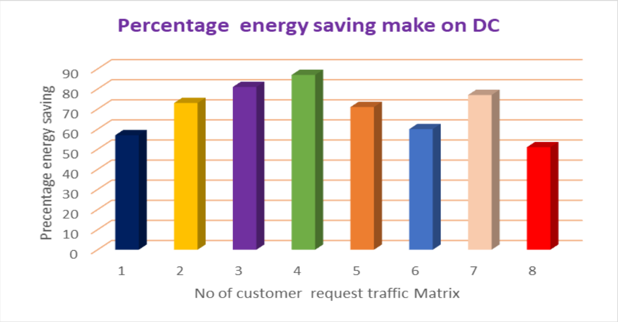
Table 4 Workload data characteristics.

Data	Number of VMs	Mean(%)	SD(%)
03/03/2011	1052	12.31	17.09
06/03/2011	898	11.44	16.83
09/03/2011	1061	10.7	15.57
22/03/2011	1516	9.26	12.78
25/03/2011	1078	10.56	14.14
03/04/2011	1463	12.39	16.55
09/04/2011	1358	11.12	15.09
11/04/2011	1233	11.56	15.07
12/04/2011	1054	11.54	15.15
20/04/2011	1033	10.43	15.21

5 Result Evaluation

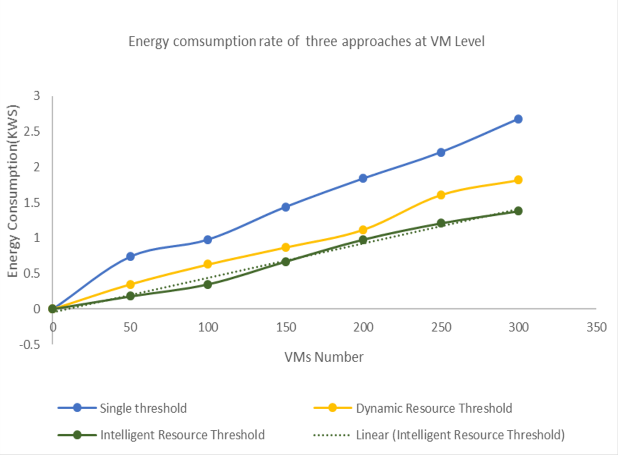
The obtained result analysis showed that this approach was able to save a significant amount of power during the processing time compared to other research works. Figure 4 shows the amount of energy saved at each virtual machine migration period using our algorithm. It is evident that at each traffic metric stage a substantial amount was saved

Figure 4: amount of energy saved at each virtual machine migration period



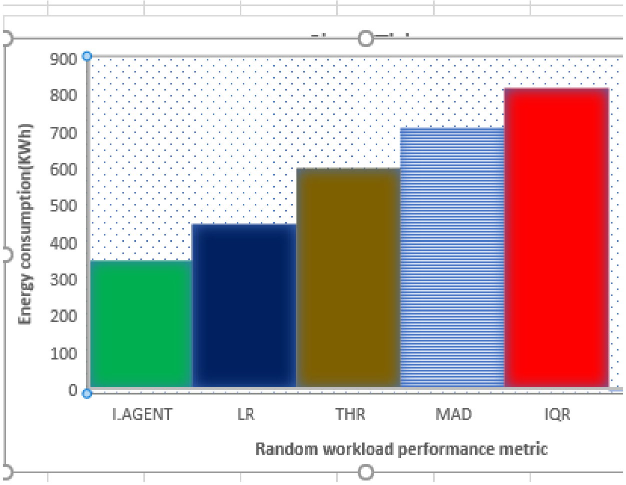
Furthermore, from the graphical result in Figure 5, it showed that when an intelligent agent code was introduced into the already existing dynamic threshold approach, it checked the migration process more efficiently and had minimal SLA violation. However, we did not include it as a graph in this paper.

Figure 5: Comparison analysis of three different VM threshold



Continuing with result evaluation, we also examined the impact of our propose algorithm to the other existing ones used on cloudsim environment which are known to be the linear regression (LR), mean absolute deviation (MAD) and interquartile range(IQR). According to Figure 6 graphical representation, it shows that the intelligent agent(I.Agent) approach consumed less power than the other scheduling approaches in the real workload traces. Figure 6 depicts that it reduced the amount of power it consumed by up to 26.9% with desirable system performance from experimental setup traces. Moreover, it achieved a significant reduction of 68.9%, 17.4%, 14.6%, and 24.1% in power consumption when compared to LR, THR, MAD, and IQR methods.

Figure 6: Power consumed by I.Agent method compared to other scheduling approaches



6 Conclusion and Future work

As cloud service providers continue to invest in innovative technology that can leverage the high-power consumption in a cloud data centre. Our proposed approach is one of its kind, that can intelligently minimise power with minimal downtime and therefore, do not encourage performance degradation. Our work brought the tradeoff between power efficiency and system performance. When compares to the existing dynamic VM consolidation methods, our proposed method provides more optimal energy-efficient solution.

Future work will model each level of SLA violation that occurred at each stage of this work. It will also test this approach on switch performance in cloudsim kit while enhancing this performance presentation.

References:

[1] Li B., et al.:EnaCloud: an energy-saving application live placement approach for cloud computing environments, Proc of international conf on cloud computing(2009)

[2] Oro E, Depoorter V, Garcia A, Salom J. Energy efficiency and renewable and energy integration in data centres. Strategies and modelling Review. Renewable and Sustainable Energy Review 2015;429-445

[3] Brown, R.: Report to congress on server and data centre energy efficiency public law 109-431," U.S. Environ. Protection Agency, Washington, DC, USA (2007).

[4] Gartner Press Release April 26, 2007, Gartner Estimates ICT Industry Accounts for 2 percent of global CO2Emissions.http://www.gartner.com/it/page.jsp?id=503867

[5] OgechukwuOkonor, Mo Adda" Intelligent Approach to Minimizing Power Consumption in Cloud-Based System Collecting Sensor data and Monitoring the Status of Powered Wheelchair" Intelligent system conference 1, 2019.

[6] OgechukwuOkonor, Mo Adda" Power Optimisation Model for Leveraging Cloud System", IEEE Conference,2019.

[7] Liu, J., Zhao, F., Liu, X., He, W.: Challenges Towards Elastic Power Management in Internet Data Centres. Proc 2nd international workshop on cyber-physical systems (WCPS), in conjunction with ICDCS Montreal, Canada, (2009

[8] Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges" Published online: The Brazilian Computer Society (2010).

[9] Tenna Mathew, KC Sekaran, J Jose –"Study and analysis of various task scheduling algorithm in the cloud computing environment". 2014 International conference …, 2014 - ieeexplore.ieee.org

[10] Nathuji R, Schwan K. Virtualpower: coordinated power Management in virtualized enterprise systems ACM SIGOPS Operating Systems Review 2007;41(6):265-278.

[11] Kusic D, Kephart J O, Hanson JE, Kandasamy N, Jiang G. Power and Performance management in virtualised computing environments via lookahead control. Cluster computing 2009,12(1):1-15.

[12] Verma A, Ahuja P, NeogiA.pMapper" power and migration cost aware application placement in virtualized systems proceedings of the 9th ACM/IFIP/USENIX international conference on middleware, Springer, Leuven, Belgium, 2008;243-264.

[13] Beloglazov," optimal Online Deterministic Algorithm and Adaptive Heuristics for Energy and performance efficient dynamic consolidation of the virtual machine in the cloud Data centre" Volume 24.

[14] ArianyaE, Taherir H, Sharifian S(2015) Novel Energy and SLA efficient resource management heuristics for consolidation of virtual machines in cloud data centres.Comput Elect Eng.

[15] Ismaeel S, Miri A(2015) Using ELM technique to predict data centre VM request. In proceeding of the 2nd IEEE international conference on cybersecurity and cloud computing(CS Cloud 2015).IEEE, New York,pp80-86.

[16] Hu B, Lei Z, Lei Y, Xu D(2011) A time series based pre-copy approach for live migration of virtual machines. In: Parallel and Distributed System(ICPADS)2011 IEEE 17th international conference on. IEEE, pp 947-952.

[17] Cui, J., Liu, S.f, Zeng, B., Xie, N.M: A novel grey forecasting model and its optimization. Applied Mathematical Modelling 37(6), 4399-4406(2013).

[18] SPEc power benchmarks, Standard Performance EvaluationCorporation. http://www.spec.org/benchmarks.html#power