Power Savings in Green Cloud Environment Using K-Means Clustering

Abhilasha Sondhi, Dr. Anupama Gupta, Dr Vivek Aggarwal

Abstract—Cloud computing is a new technology which provides data access, storage services and computation as services to the consumers through the Internet, based on the pay-as-you-go model. Green cloud computing spawned to tackle large amount of energy consumption and the rise in carbon dioxide emission. In the present work, there are three processors running on three different frequencies namely maximum frequency of 100%, DVFS (Dynamic voltage and frequency scaling) and threshold frequency which handle the various real time and non-real time cloudlets. The cloudlets are clustered using the k-means algorithm where the Euclidian distance is calculated for the three parameters which are the instruction size of cloudlet, the deadline of cloudlet and the cost paid by the customer. Thus, the priority is decided by clustering of the incoming cloudlets on the basis of these three parameters. After quantitative analysis, significant improvement is noticed over the existing system related to power consumption, total processing time, total turnaround time, waiting time and total processing cost.

Index Terms— Green cloud computing, power consumption, dynamic voltage frequency and scheduling, data centres,virtual machines,cloudlets.

1 Introduction

Cloud Computing is a new paradigm which delivers computing as a utility through the internet. It provides data access, software, storage services and computation as services which are provided to the consumers through the Internet based on the pay-as-you-go model. It provides substantial benefits to the IT companies as they are comforted from making the setups for hardware and software infrastructure and thus decreasing the cost of those companies (Wadhwa and Verma, 2014). Internet has been a mainspring towards the numerous technologies that have been expanded since its commencement.

Green cloud computing is becoming important due to the increasing concerns about environmental issues by the large Internet and cloud service providers. However, with the growing demands of users for computing services, cloud providers are encouraged to use big data centers, which consume large amounts of energy resulting in carbon dioxide emissions. Power consumption is a key concern in data centers. Such type of critical issues not only reduces the profit margin, but also has effect on high carbon production, which is harmful for the environment and living organisms. Reducing power consumption has been an essential need for cloud resource suppliers not only to reduce operating costs, but also to improve system reliability. A number of practices can be applied to achieve energy efficiency, such as enhancement of algorithms, energy proficient hardware, Dynamic Voltage and Frequency Scaling (DVFS), terminal servers and thin clients, and virtualization of computer resources. Dynamic voltage scaling (DVS) has been a key technique in exploiting the hardware features of cloud data centers to protect energy by lowering the supply voltage and operating frequency. The main goal is to reduce the power consumption by data centers. Energy efficient scheduling of workloads helps to reduce the consumption of energy in datacenters, thus helps in better usage of resources. This further reduces operational costs and provides benefits to the consumers and service providers.

2 Literature Review

Buyya et al. (2011) presented open research challenges, algorithms for management of energy in Cloud computing environments. The proposed energy-aware allocation heuristics provide data center resources to customer applications in a manner that increases energy efficiency of the data center, though distributing the negotiated Quality of Service (QoS). They led a survey of research and proposed: (a) architectural principles for energy-efficient organization of Clouds; (b) resource allocation policies and scheduling algorithms considering QoS expectations and power usage features of the devices; and (c) several open research challenges, addressing which can bring significant benefits to both resource providers and consumers.

Reddy and HKS (2014) have presented a system that executes real-time and non-real-time jobs in an energy efficient method. They used three processors, out of which i.e. the first and second, execute real-time tasks, using earliest-Deadline-First (EDF) and Earliest-Deadline-Late (EDL) scheduling algorithms respectively. The third processor, handles non-real-time tasks using the First Come First Served (FCFS) scheduling algorithm .The simulation results showed significant energy savings compared to the existing Stand-by Sparing for Periodic Tasks (SSPT) for some execution set-ups.

Lei et al. (2014) proposed a smart green energy-efficient scheduling approach for the task scheduling in the data center, based on the renewable energy prediction and the dynamic grid electricity price. The projected strategy efficiently achieved a low total cost of energy consumption, and a high utilization rate of renewable energy, keeping a good rate of task satisfaction for the task scheduling according to the prediction of renewable generation and the dynamic grid electricity expense.

Roy and Gupta (2014) have proposed framework which enables energy efficient technique of a minimal discharge and refinement of the problem of great carbon production so that the profit margin increases. The profitability was expressed in terms of bio and environment friendliness rather than with respect to revenue. The target was to merge the cloud service with the green cloud computational framework for the future research.

Wadhwa and Verma (2014) have proposed a new technique to decrease the carbon emission and energy consumption in the distributed cloud data centers having different energy sources and carbon footprint rates. Carbon footprint rate of the data centers in distributed cloud architecture had been used and the concept of virtual machine provisioning and migration for dropping the carbon emission and energy consumption in the federated cloud system. Simulation results indicated that the proposed method lowered the carbon dioxide emission and energy consumption of federated cloud datacenters as matched to the conventional scheduling method of round robin Virtual machine (VM) scheduling in federated cloud datacenters.

Yakhchi et al. (2015) have presented an approach based on Cuckoo Optimization Algorithm (COA) to identify over-utilized hosts. The Minimum Migration Time (MMT) policy was used to migrate Virtual Machines (VMs) from the over-utilized hosts to the under-utilized hosts. The results generated by Cloudsim simulator, confirmed that the proposed approach has the lowest energy consumption compared to the other well-known algorithms.

Xia et al. (2015) have presented a novel stochastic framework for energy efficiency and performance analysis of DVS-enabled cloud. This framework made use of virtual machine request arrival rate, failure rate, repair rate, and service rate of datacenter servers as model inputs. Dependent on a queuing network-based analysis, the paper gave analytic explanations of the three metrics. The proposed framework can be used to support the design and optimization of energy-aware high performance cloud systems.

Nagpure et al. (2015) have proposed dynamic resource allocation system which allocated resources to cloud user. The skewness algorithm was applied to measure irregular use of multiple resources of each VMs and hence balanced across VMs. By reducing the skew value of each VM, the client could simply combine different multiple resources and enhance resource utilization of servers These proposed algorithms stopped overloading by effective load balancing and future load prediction and attained improved performance, of server resource utilization by minimum energy consumption.

Hasan et al. (2015) have presented a scheme for green energy management in combination of renewable energy in data center. They proposed three contributions: i) to address the uncertainty of green energy availability, they introduced the concept of virtualization of green energy ii) to support Green SLA they introduced two new threshold parameters and extended the Cloud Service Level Agreement (CSLA) language and iii) presented green SLA algorithm which influences the concept of virtualization of green energy to deliver per interval specific Green SLA. PlanetLab conducted experiments with real workload profile from and server power model from SPECpower to illustrate that, Green SLA could be effectively established and satisfied without incurring higher cost.

Chen et al. (2015) have focused on ''greening'' demand response in multi-tenant data centers by designing a pricing mechanism through which the data center operator could efficiently extract load reductions from tenants during emergency periods for EDR (Emergency Demand Response). In particular, they proposed a pricing mechanism for both mandatory and voluntary EDR programs, ColoEDR, based on parameterized supply function bidding and provided provably near-optimal efficiency guarantees, both when tenants were price-taking and when they were price-anticipating. In addition to analytic results, they extended the literature on supply function mechanism design, and evaluated ColoEDR using trace-based simulation studies. These validated the efficiency analysis and concluded that the pricing mechanism was both beneficial to the environment and to the data center operator (by decreasing the need for backup diesel generation), while also aiding tenants (by providing payments for load reductions).

Paul et al. (2015) presented that, the proposed technique of modelling switching cost through variance achieved a better trade-off between some important parameters instead of modelling switching cost through a linear function Since the three major input parameters-electricity price, renewable energy

ISSN 2229-5518

and number of job requests-vary overtime, the average cost of electricity per job request may also exhibit dramatic fluctuations. They proposed to tackle the volatility by controlling the average cost of electricity per job request through leveraging contracts in the forward electricity market, and determined the optimal amount of electricity to be procured in the forward electricity market.

3 Proposed Methodology

The proposed work puts forward an efficient energy consumption technique, keeping in mind the problems formulated in the existing system. In proposed technique cloud environment is developed in java, deployed on CloudSim toolkit. The CloudSim simulator is initialized using the appropriate number of cloudlets, in the present case these cloudlets range from ten thousand to ninety thousand and they run on the virtual machines in datacenter. Then, the classification of the cloudlets is done into different categories depending upon the nature of the tasks. k-means clustering algorithm is used to divide the cloudlets into multiple clusters depending upon cloudlet's length, deadline and cost. First cluster contains the cloudlets of higher instruction size and earlier deadline similarly the second cluster contains the cloudlets with medium instruction size and deadline greater than the cloudlets in first cluster and finally the third Cluster contains the cloudlets with least instruction size and are of lesser priority. Broker receives all the cloudlets from the user and the virtual machines from the datacentre. Broker finally applies the power efficient load balancing algorithm in which the cloudlets of Cluster 1 is be assigned to first processor (P1). The P1 processor executes all the cloudlets of Cluster 1 at the maximum frequency .Cloudlets available inside the Cluster 2 are be assigned to second processor (P2). It executes at the DVFS mode. Dynamic voltage scaling is a technique to manage power utilization of processors. Dynamic refers to increasing or decreasing voltage which depends upon circumstances The processor P2 checks whether the assigned cloudlet of Cluster 2 are executed at the full frequency or at the threshold mode. The third processor (P3) runs all the cloudlets that are in the third cluster at the threshold frequency.

The experimental results have been compiled as per quantitative analysis. In proposed technique , power savings in green cloud environment has been done using k-means clustering to sort the cloudlets according to the length of instruction, their deadline and price paid by the user so that the cloudlet can be segregated as per their priority and allocated to processor to run at either at full frequency or DVFS. The Figure1 depicts the schematic of the proposed work.

3.1 Use of k-means clustering algorithm

The process of grouping a set of entities into classes of similar entities is the central use of K-means clustering. Clusters are based on centroids of points in a cluster, *c*. Relocation of instances to clusters is dependent on distance to the current cluster centroids.

In the present work, cloudlets are divided into the clusters according to the cloudlet's length, priority and cost. Euclidean distance is computed for finding the distances of the Cloudlet with the Cluster's centroid.

$$\text{Euclidean Distance} = \sqrt{(\text{Length}_i - \text{Length}_j)^2 + (\text{Cost}_i - \text{Cost}_j)^2 + (\text{Priority}_i - \text{Priority}_j)^2} \quad (1)$$

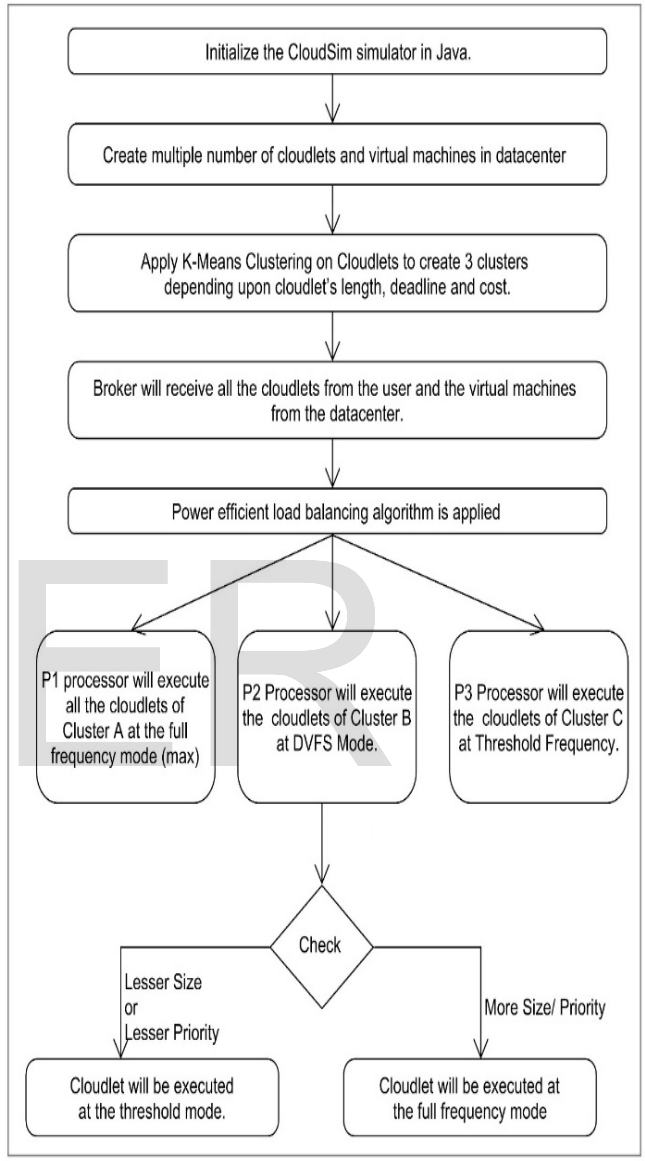

Fig 1 The schematic of proposed methodology

3.1 Simulation Results and Discussion

A number of experiments were conducted by creating number of cloudlets virtually in the cloudsim environment. Various quantitative metrics such as power consumption, processing cost, processing time, waiting time and turnaround time were computed for both base and proposed system. The graphical results corresponding to the metrics have been shown in Figures 2 -6. These results show the comparisons when experiments were conducted on the base system and the proposed system.

Power Consumed: It is defined as the product of voltage and total time taken by the processor to finish the execution of a cloudlet.

Power consumed = Voltage*time (2)

Processing Cost: It is obtained by multiplying the unit cost of electricity with processing time of a cloudlet.

Processing Cost = processing time * unit_cost. (3)

Processing Time: It is defined as total time taken to complete the execution of a cloudlet.

Processing Time = finish time – start time (4)

Turnaround Time: It is the interval from time of submission of the process to the time of completion of the process. Turnaround Time = Finish Time -Arrival Time (5)

Waiting time: It is the amount of time the process uses the processor before it is ready.

Waiting time=Start time-Arrival time (6)

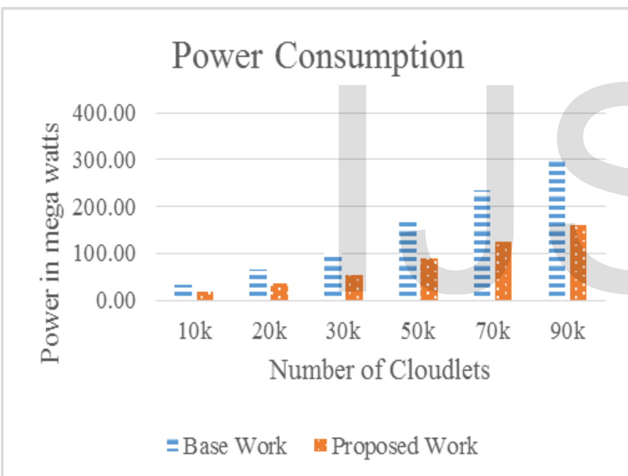

Fig 2 The bar graph of power consumption

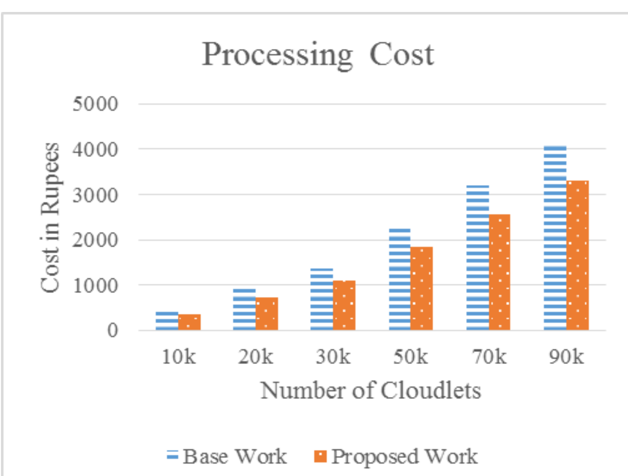

Fig 3 The bar chart of processing cost

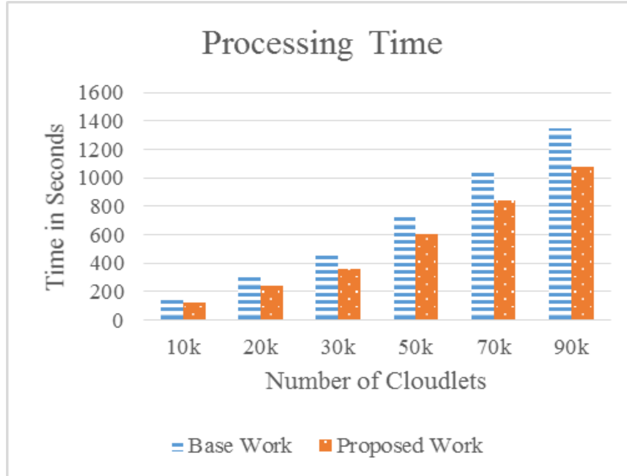

Fig 4 The bar graph of processing time

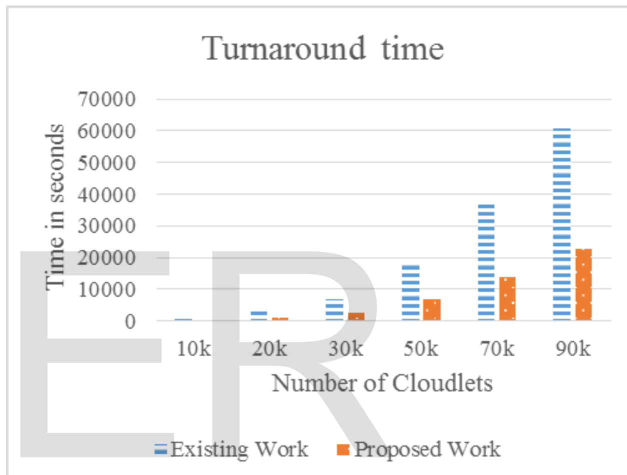

Fig 5 The bar graph of Turnaround Time

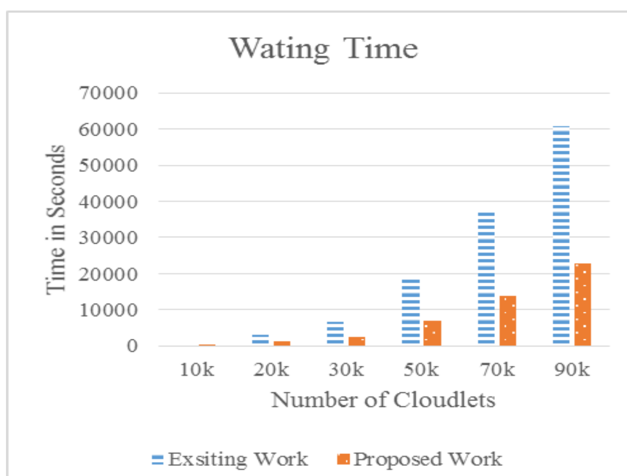

Fig 6 The bar graph of waiting time

4 Conclusions and Future Scope

The proposed technique shows 46% reduction in power consumption, 62% reduction in waiting time and turnaround time and 20% reduction in both processing time as well as cost paid by the user. The advantage of reducing power consumption is that less heat is generated by device which benefits the mechanical design as well as it improves the lifetime period of devices. In addition, low power consumption by the infrastructure leads to drop in carbon dioxide (CO2) emissions contributing to the concept of green cloud computing. In the future work can be extended for achieve Green Cloud Computing server data centers can use renewable energy sources like the solar systems, bio-gas plant energy, wind energy can provide power to data centers as these are eco-friendly sources of energy.

References

[1] Buyya, R., Beloglazov, A., Abawajy, J (2011).” Energy- Efficient Management of Data Center Resources for Cloud Computing: A Vision, Architectural Elements, and Open Challenges", IEEE.

[2] Reddy, S. P., & H K S, C. (2014)." Energy Aware Scheduling of Real-Time and Non Real-Time Tasks on Cloud Processors", IEEE.

[3] Wadhwa, B., & Verma, A. (2014). "Carbon Efficient VM Placement and Migration Technique for Green Federated Cloud Datacenters", Chandigarh,India: IEEE.

[4] Roy, S., & Gupta, S. (2014)." The Green Cloud Effective Framework: An Environment Friendly Approach Reducing CO2 Level". Kalyani, India: IEEE.

[5] Nagpure, B. M., Dahiwale, P., & Marbate, P. (2015). "An Efficient Dynamic Resource Allocation Strategy for VM Environment in Cloud". Nagpur, India: IEEE.

[6] Lei, H., Zhang, T., Liu, Y., Zha, Y., & Zhu, X. (2015). "SGEESS: Smart green energy-efficient scheduling strategy with dynamic electricity price for data center". The Journal of Systems and Software (pp. 23–38). Changsha, Hunan 410073, PR China: Elsevier.

[7] Xia, Y., Zhou, M., Luo, X., Pang, S., & Zhu, Q. (2015)." A Stochastic Approach to Analysis of Energy-Aware DVS-Enabled Cloud Datacenters". IEEE.

[8] Yakhchi, M., Ghafari, S. M., Yakhchi, S., Fazeli, M., & Patooghi, A. (2015). "Proposing a Load Balancing Method Based on Cuckoo Optimization Algorithm for Energy Management in Cloud Computing Infrastructures", IEEE.

[9] Ariharan, V., & Manakatu, S. S. (2015). "Neighbour Aware Random Sampling (NARS) algorithm for load balancing in Cloud computing". Bengaluru,India: IEEE .

[10] Dehouche, N. (2015)." A Bi-criteria Algorithm for Low-Carbon and QoS-Aware Routing in Cloud Computing Infrastructures". Northern Borders University: IEEE.

[11] Fahimeh Farahnakian, T. P. (2015)." Utilization Prediction Aware VM Consolidation Approach for Green Cloud Computing". Turku, Finland: IEEE 8th International Conference on Cloud Computing.

[12] Giacobbe, M., Celesti, A., Fazio, M., Villari, M., & Puliafito, A. (2015)." An Approach to Reduce Carbon Dioxide Emissions through Virtual Machine Migrations in a Sustainable Cloud Federation". Messina, Italy: IFIP.

[13] Goiri, Í, Haque, M. E., Le, K., Beauchea, R., Nguyen, T., Guitart, J.Bianchini, R. (10 November 2014). "Matching renewable energy supply and demand in green datacenters". Rutgers University, USA: Elsevier

[14] Hasan, M. S., Kouk, Y., Ledoux, T., & Pazat, J. L. (2015). "Exploiting Renewable sources: when Green SLA becomes a possible reality in Cloud computing". INSA de Rennes, France. : IEEE.

[15] Ketankumar, D. C., Verma, G., & Chandrasekaran, K. (2015). "A Green Mechanism Design Approach to Automate Resource Procurement in Cloud". Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015). Mangalore 575 025, India: Elsevier.

[16] Oró, E., Depoorter, V., Garcia, A., & Salom, J. (15 October 2014)." Energy efficiency and renewable energy integration in data centres. Strategies and modelling review". Catalonia Institute for Energy Research, IREC, Jardins de les Dones de Negre 1, 08930 Sant Adrià de Besòs, Barcelona, Spain: Elsevier.

[17] Peng, Y., Kang, D. -K., Hazemi, F. A., & Youn , C. -H. (19 February 2016). "Energy and QoS aware resource allocation for heterogeneous sustainable cloud datacenters". Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea: Elsevier.

[18] Qiu, C., Shen, H., & Chen, L. (2015). "Towards Green Cloud Computing: Demand Allocation and Pricing Policies for Cloud Service Brokerage". International Conference on Big Data (Big Data). Clemson, USA: IEEE.

[19] Hasan, M. S., Kouk, Y., Ledoux, T., & Pazat, J. L. (2015). "Exploiting Renewable sources: when Green SLA becomes a possible reality in Cloud computing". INSA de Rennes, France. : IEEE.

[20] Chen, N., Ren , X., Ren , S., & Wierman , A. (2015). "Greening multi-tenant data center demand response". California Institute of Technology, United States: Elsevier.

[21] Paul, D., Zhong, W, Bose,S.(2015). ‘Energy efficiency aware load distribution and electricity cost volatility control for cloud service providers’. Elsevier.