

# An Approach for Resource Management in Cloud Computing to Reducing Power Consumption

Bhushan A.Ugale  
Dept. Of Information  
Technology, M.C.O.E.R.C,  
Nashik,India  
[bhushan.ugale@gmail.com](mailto:bhushan.ugale@gmail.com)

Sahil Sakhala  
Dept. of Computer  
Engineering, K.K.W.I.E.E.R  
Nasik, India  
[sahils1210@gmail.com](mailto:sahils1210@gmail.com)

Jigneshkumar H. Patel  
Dept. Of Information  
Technology,TITS,  
Modasa,India  
[jignesh.dholu@gmail.com](mailto:jignesh.dholu@gmail.com)

**Abstract**— Resource Management and Power consumption is one of the most critical problems of data centres. One effective way to reduce power consumption is to consolidate the hosting workloads by effective way of Resource management and reducing power consumption by using the computation resource in such a way that they can give optimal output by shut-down physical machines which become idle after consolidation[1] and by using the effective algorithm to use the resources in a better manner. In this paper we try to reduce the power consumption of the computational machine used in educational sector by using effective resource management technique and also use the concept of “Time-of-use pricing” to reduce the computation cost of the cloud.

**Keywords**— DVFS, VM, DA, SLA, SETI, DPM, PA-LRU, PB-LRU.

## I. INTRODUCTION

According to the EPA report, servers and data centre consumed about 61 billion kilowatt-hours (kW·h), and it was 1.5% of the total U.S. electricity consumption for about \$4.5 billion electricity cost in 2006. The power consumption still keep growing at 18% annually, and it also estimates that the power consumption of data centre could nearly double by 2011.[2] Using effective algorithms we can reduce the computational cost and power required to do the same. In this paper we focused on some algorithms and technique like redundancy avoidance, Gale-Shapley algorithm, Job Machine Stable Matching and technique like Dynamic Voltage Frequency Scaling (DVFS). We also give our proposed model which is based on the Educational Institutions where most of the computational machine stay under utilised and we can use there un-utilised processor to gives power to the machine used in villages for general purpose computing.

## II. CLOUD COMPUTING AND CHALLENGES

Due to the multi-tenant nature, resource management becomes a major challenge for the cloud. According to a 2010 survey, it is the second most concerned problem that CTOs express after security. Cloud operators have a variety of distinct resource management objectives to achieve.[3] In a cloud computing environment, resources are shared among different clients. Intelligently managing and allocating resources among various clients is important for system providers, whose business model relies on managing the infrastructure resources in a cost-effective manner while satisfying the client service level agreements (SLAs).[4]and also considering the cost against efficiency to make the cloud computing more effective and efficient we have to consider the better resource management practice.

### A. RESOURCE MANAGEMENT BY STABLE MATCHING ALGORITHM.

It works by having agents on one side of the market, say men, propose to the other side, in order of their preferences. As long as there exists a man who is free and has not yet proposed to every woman in his preference, he proposes to the most preferred woman who has not yet rejected him. The woman, if free, “holds” the proposal instead of directly accepting it. In case she already has a proposal at hand, she rejects the less preferred. This continues until no proposal can be made, at which point the algorithm stops and matches each woman to the man (if any) whose proposal she is holding. The woman-proposing version works in the same way by swapping the roles of man and woman. It can be readily seen that the order in which men propose is immaterial to the outcome. Resource management in the cloud can be naturally cast as a stable matching problem, where the

overall pattern of common and conflicting interests between stakeholders can be resolved by confining our attention to outcomes that are stable. Broadly, it can be modelled as a many-to-one problem where one server can en-roll multiple VMs but one VM can only be assigned to one server.. A large body of research in economics has examined variants of stable matching and references therein. Algorithmic aspects of stable matching have also been studied in computer science. However, the use of stable matching in networking is fairly limited. uses the DA algorithm to solve the coupled placement of VMs in data-centres. [3] this algorithms considers the client interest and shut the machine off which is less preferred and migrate the VM to other one which is most preferred.

### B. RESOURCE MANAGEMENT BY AVOIDING REDUNDANCY.

Backup services require data to be backed up to dedicated external drives, which can be inconvenient or costly to the users, data backup for personal computing has emerged to be a particularly attractive application for outsourcing to the cloud storage providers because users can manage data much more easily without having to worry about maintaining the backup infrastructure. In source deduplication, elimination of duplicate data occurs close to where data is created, rather than where data is stored as in the case of target deduplication. Performing deduplication at the source can dramatically improve IT economics by minimizing storage requirements and network bandwidth consumption since the redundant data is eliminated prior to its traverse across the network to the target backup server. Data chunking has a significant impact on the efficiency of deduplication. In general, the

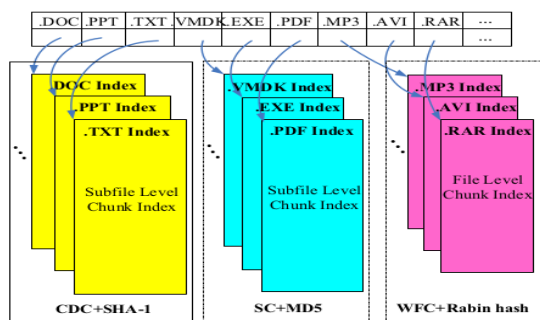


Fig.1 Application-Aware index structure

Deduplication ratio is inversely proportional to the average chunk size. On the other hand, the average chunk size is also inversely proportional to the space overhead due to file metadata and chunk index in storage systems. In order to determine which chunks have already been stored. If there is a match in the index, the incoming chunk contains redundant data and can be deduplicated to avoid transmitting it; if not, the chunk needs to be added to the cloud storage and its hash and metadata inserted into the index. [5].

Deduplication effectiveness is very important for both cloud backup providers and users. Providers expect less data stored in their data centre to reduce data storage and management costs, whereas users prefer transferring less data for shorter backup time and lower storage cost. 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. [6] and taking better outcomes from the virtualization we need to use the virtualization in optimized ways.

### C.PERFORMANCE & RELIABILITY OF VIRTUALIZATION

Virtualized cloud resources enable performance isolation, on-demand creation and customization of execution environments, from which scientists can benefit. The goal is to improve energy efficiency, while maximizing the rate of completed jobs (i.e. SLA), by constructing reliable and high-available computing environments. The recently resurgence of virtualization relates to the possibility to deal with issues in the data-centre, such as hardware underutilization, variance in demanding requirements of computing resources, high system administration costs, and reliability and high-availability.[7]

### III.POWER SAVING BY FREQUENCY SCALING

Traditionally, systems have been designed to achieve maximum performance for the workload, but report from NRDC indicates that servers still use 69-97% of total energy when idle. Therefore, we can avoid waste and reduce power consumption by scaling down CPU frequencies and changing throttling levels via DVFS to minimize the processor power dissipation. DVFS is a technique in computer architecture whereby the frequency of a microprocessor can be automatically adjusted "on-the-fly" to conserve power and to reduce the amount of heat generated by the chip. With the hardware support of Intel® SpeedStep and AMD PowerNow!™ technologies, modern processors can be set in all available frequencies and throttling levels on different voltage. Hence, we adopt it into this cloud architecture to reduce processor power dissipation, and set demand frequencies and throttling levels at idle time or application runtime on different phases. One of the open source tool named cpufreqd is used to help users to set demand frequencies for dedicated applications automatically and lower down after finishing jobs according to the rules defined by users.[2] Also we can migrate the virtual machine to other physical machine to make a machine idle which is underutilized. So that if one of the machine underutilised and other can able to take load of the same in off peak time then we can save the power in a better manner. Dynamic resource placement and provisioning are useful techniques for handling the

multi-time-scales variations in the capacity of resources. Supported by making decision algorithms, dynamic resources allocation is used to perform consolidation targeting goals such as minimizing total power consumption, without substantial degradation on performance. Concerning resource consolidation and optimization of power consumption. A dynamic consolidation resource manager for homogeneous clusters based on constraint programming that takes migration overhead into account. The goal is to minimize the number of active nodes, and powering down inactive ones, while maintaining performance. The VMs were considered to be active or inactive.[7]

Live migration allows a running virtual machine or application to move between two physical servers without disconnecting the client or application. The power overhead incurred by live migration is very small, which also has been verified by our experiments. Power consumption of an idle CPU is half of the peak power consumed by a fully loaded CPU. This power model has been experimentally verified with real VM workload execution on an eight-node cluster and with actual power consumption measured by power meters. For the migration model, live migration is used in our work.

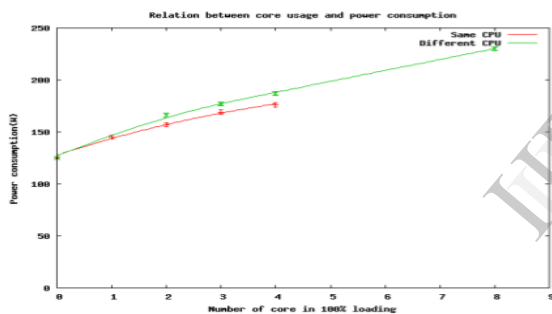


Fig. 2 Power Consumption Under Different CPU Loads

Live migration allows a running virtual machine or application to move between two physical servers without disconnecting the client or application. The power overhead incurred by live migration is very small, which also has been verified by our experiments.[1]

#### A. AN ALGORITHMIC APPROACH

Dynamic Round-Robin method is proposed as an extension to the Round-Robin method. Dynamic Round-Robin method uses two rules to help consolidate virtual machines. The first rule is that if a virtual machine has finished and there are still other virtual machines hosted on the same physical machine, this physical machine will accept no more new virtual machine. Such physical machines are referred to as being in “retiring” state, meaning that

when the rest of the virtual machines finish their execution, this physical machine can be shutdown.

Reducing power consumption has been an essential requirement for Cloud resource providers not only to decrease operating costs, but also to improve the system reliability. Most resource-intensive companies growing demand for high performance computing infrastructures.

#### B. DATACENTRE LOAD MANAGEMENT

$$load(h) = \frac{\mathcal{N}(h)}{\mathcal{C}(h)}$$

The ability to dynamically adjust the performance of computer components proportionally to their power consumption is called Dynamic Performance Scaling (DPS). It is possible to adjust the computer supply voltage when it is not fully utilized. Based on this idea, many techniques are adopted. In this algorithm, the nominal MIPS ( $\mathcal{N}$ ) for each host represents the maximum computing capability of the host at the maximum frequency, while the host load ( $\mathcal{C}$ ) represents the current load of the host in MIPS. The load on host  $h$  is computed using equation below, which is equal to the ratio of maximum computing

$$load(DC) = \sum_{\forall h \in \mathcal{H}} \frac{\mathcal{N}(h)}{\mathcal{C}(h)} / Size(\mathcal{H})$$

capability to current load. The data-centre load is computed using below equation, which is equal to the average load on all its hosts.

Using this output we can make decision for which algorithms and technique we have to use to fulfill out requirement.[8]

#### V. PROPOSED STRATEGY FOR EDUCATIONAL INSTITUTION

In our proposed strategy we consider the scenario of the educational institution. In our case (Engineering college) where we observe the CPU and DISK utilization of the computers and found most of the time there utilization of the CPU and DISK are very less. When that machines are underutilised we can use there processor and disk to give power to social cloud where we can offer small software like spread sheet and small amount of disk storage to give advantage of computing to the small business of the village to enhance there computing. All the machine within this cloud are volunteer machine that mean they are not bound to any SLA but can give there underutilised power to the required one. Most of the time some machines of the institute are in idle mode. So that they can give computing power to this social cloud.

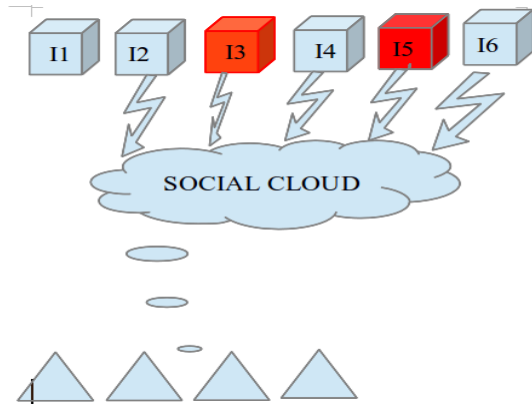


Fig. 3 Social cloud based on volunteer Processors

In the above figure I1 to I6 represents institutional machine which gives their unutilised power to the social cloud and triangle represent the thin client with lower power consumption or the beneficiary machine. Which can be used for the small scale business. We are using the concept relevant to the SETI(Search for Extra Terrestrial Intelligent) where they use the power for volunteer machine for their calculation. In our case we are doing so to give the processing power to our social cloud.

#### VI. COST SAVING BY MIGRATING VIRTUAL MACHINE.

The virtual machine gives the freedom to migrate the machine from any physical machine to other physical machine as per the requirement so we can use the concept of "Time-of-use pricing" which is a rate structure that reflects the costs associated with electricity production throughout the day. Prices rise and fall over the course of the day and tend to drop overnight and on weekends, depending on demand and the availability of supply.[9] To get the benefits of the same concept we can migrate our virtual machine dynamically to the server which are in the off-peak zone and can get the benefits of the reduced cost. Which ultimately makes our computation cost effective.

#### VII. POWER-SAVING BY CACHING AND BUFFERING OF MEMORY

Caching and buffering in memory saves energy. To improve performance, data blocks read from disk, are cached in memory, because memory is several orders of magnitude faster than a disk and recently-read blocks have a high probability to be read again soon after. This workload characteristic is called temporal locality. For every cache hit, a disk access is avoided. This means that caching results in longer disk idle times, facilitating DPM(Dynamic Power Management) Modified data blocks, which need to be written to disk for persistent storage, are buffered in memory such that the application performance is not

impacted by the large disk response time. When a buffered block is modified again, a disk access is avoided.

This increases the time the disk is idle, creating more opportunities for saving power using DPM [10].

#### VIII. POWER SAVING BY CACHE REPLACEMENT POLICY

The storage-cache replacement policy determines which data block is replaced when a cache miss occurs. It is designed for optimal performance by minimizing the number of disk accesses Here we will discuss two methods power-aware LRU (PA-LRU) and partition-based LRU (PB-LRU)

PA-LRU classifies the disks in the array in two categories, as shown in Figure 4:

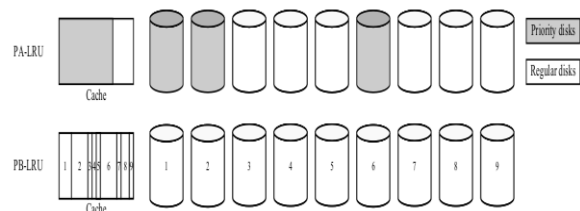


Fig.4 Power-Aware cache replacement policies :PA-LRU & PB-LRU

regular and priority. This cache replacement algorithm prioritizes disks that exhibit a large percentage of long idle periods and a high ratio of capacity misses to cold misses. PA-LRU selects the least-recently used regular-disk block for eviction. Only when no blocks from regular disks reside in the cache, the least-recently used priority-disk block is evicted. Thus, the cache tends to hold more blocks from priority disks than from regular disks. As a consequence, PA-LRU directs I/O requests mostly towards the regular disks. This means that the priority disks may increase their mean idle time, while for the regular disks the idle time is reduced, on average. The higher idle time variance results in larger energy savings. [10]

#### IX .POWER SAVING BY OPTIMIZING DISK USES

Fig. 1 shows the percentage of power consumed by typical system components as a percentage of the overall power consumption. Power consumption is dominated by the storage subsystem. The greatest portion of all power, 79.1%, is consumed by the storage subsystem, i.e. disks, controllers and disks. Disks are the largest consumer at 63%. Client and server systems consume 16.9%, of which 8.6 percent is used by memory and 5.8 percent is consumed by the CPUs.

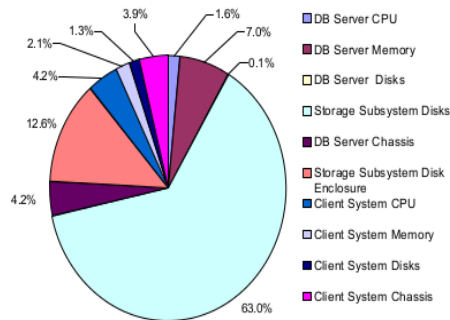
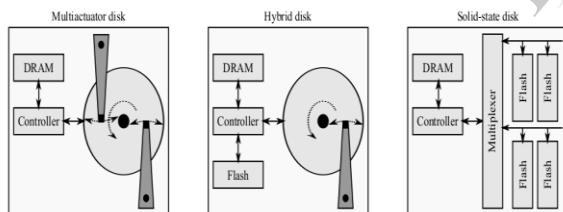


Fig.5 Average power consumption of major Components

So by making the disk more power efficient is a right way to reduce the power consumption.[11] To achieve this goal we can approach the technique discussed in section II(B).Which can reduce the extra uses of the disk. The number of disks in a high-performance storage system is determined by requirements of performance rather than capacity. This leads to an underutilization of the available disk space and, thus, a waste of energy. Improve disk performance such that disks can be completely filled and, hence, a minimum number of disks is required. By simplifying the design of a multiactuator disk by not allowing more than one access arm to move simultaneously and by letting only one head at a time transfer data. These restrictions ensure that such multiactuator disk has a peak power consumption comparable to a conventional disk also by using Hybrid disks. Which is the combination of a conventional disk with NAND flash memory, which serves as a second-level cache,



as shown in Figure 6

Fig. 6 Energy-efficient disk drive architecture

NAND flash memory. NAND flash memory is non-volatile memory that is accessed in a similar way as a block device, such as a disk. A NAND flash memory device contains in a single package (or channel) multiple dies (also called chips or ways)[10].

## X. CONCLUSION

Reducing power consumption has been an essential requirement for Cloud resource providers not only to decrease operating costs, but also to improve the system reliability. Most resource-intensive companies growing demand for high performance computing

infrastructures. In this paper we try to focused on the different technique which can used to save power of the cloud computing environment and also discussed about the effective algorithms which can be use to utilised the available resource in a optimised way. Also we present our proposed model which uses the under-utilised computing power of the institutional machine which are in idle mode.

## XI. REFERENCES

- [1] Ching-Chi Lin,Pangfeng Liu,Jan-Jan Wu,"Energy-efficient Virtual Machine Provision Algorithms for Cloud Systems",2011 Fourth IEEE International Conference on Utility and Cloud Computing.
- [2] Che-Yuan Tu, Wen-Chieh Kuo, Steven Shiau ,"A Power-Aware Cloud Architecture with Smart Metering ",2010 39th International Conference on Parallel Processing Workshops .
- [3] Hong Xu,Baochun Li, "Anchor: A Versatile and Efficient Framework for Resource Management in the Cloud ",IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, 201X.
- [4] Pengcheng Xiong, Yun Chi, Shenghuo Zhu, Hyun Jin Moon, Calton Pu, Hakan Haciguzmıs, "Intelligent Management of Virtualized Resources for Database Systems in Cloud Environment",ICDE Conference 2011.
- [5] Yinjin F , Hong Jian,Nong Xiao ,Lei Tian,Fang Liu,"AA-Dedupe: An Application-Aware Source Deduplication Approach for Cloud Backup Services in the Personal Computing Environment " ,2011 IEEE International Conference on Cluster Computing.
- [6] Liang-Teh Lee, Kang-Yuan Liu, Hui-Yang Huang and Chia-Ying Tseng , "A Dynamic Resource Management with Energy Saving Mechanism for Supporting Cloud Computing " ,International Journal of Grid and Distributed Computing Vol. 6, No. 1, February, 2013.
- [7] Altino M. Sampaio,Jorge G. Barbosa ,"Dynamic Power-and Failure-Aware Cloud Resources Allocation for Sets of Independent Tasks ",2013 IEEE International Conference on Cloud Engineering.
- [8] Nawfal A. Mehdi, Ali Mamat, Ali Amer, Ziyad.T. Abdul-Mehdi , "Minimum Completion Time for Power-Aware Scheduling in Cloud Computing " ,2011 Developments in E-systems Engineering.
- [9] <http://www.energy.gov.on.ca/en/smart-meters-and-tou-prices/>
- [10] Tom Bostoen, Sape Mullender ,Yolande Berbers, "Power-Reduction Techniques for Data-Center Storage Systems" ,ACM Computing Surveys, Vol. 45, No. 3.
- [11] Meikel Poess, Raghunath Othayoth Nambiar, "Tuning Servers, Storage and Database for Energy Efficient Data Warehouses" ,