ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

New balancing technique for green cloud computing and environmental Sustainability

Dr. Ayman E. Khedr, Dr. Mona Nasr, Hesham Elmasry
Faculty of Computers & Information Helwan University

Manuscript Info

Manuscript History:

Received: 22 July 2015
Final Accepted: 25 August 2015
Published Online: September 2015

Key words:

Load balancing, green cloud computing, energy Efficiency, quality of service and data center

***Corresponding Author**

Dr. Ayman E. Khedr

Abstract

The growing demand of Cloud Computing Infrastructure has increased the energy consumption of data centers, which has become a critical issue.
Data centers hosting cloud computing applications which consume huge amounts of energy, contributing to high operational costs and leads to high carbon emissions which are not environmentally friendly. Therefore we need to propose a green cloud load balancing (GCLB) solution that intended to reduce energy consumption in cloud data center while maintaining the service level agreement (SLA) between the customer and the cloud service provider.
in order to design such solutions, deep analysis of Cloud is required with respect to their power efficiency. Thus, in this paper, we discuss various elements of green clouds which contribute to the total energy consumption and how it is addressed in the previous studies. We also discuss the implication of this solution on the energy efficiency (EE) and the quality of service (QOS).

INTRODUCTION

The energy-efficient load balancing techniques on cloud computing aims to maintains and enhances both quality of service (QOS) and resource utilization while reducing the power consumption to achieve the environmental sustainability concept.

The generally definition of cloud computing (CC) comes from The National Institute of Standards and technology (NIST) as follows: "cloud computing is a model for enabling convenient, on-demand network access To a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction." [1]

NIST also offers up several characteristics that it sees as essential for a service to be considered "Cloud". These characteristics include;

- *On-demand self-service:* the ability for an end user to sign up and receive services without the long delays that have characterized in traditional IT.

- *Broad network access:* the ability to access the service through the standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile Phones, laptops, and PDAs).

- *Resource pooling:* The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided Resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, Network bandwidth, and virtual machines.

- *Rapid elasticity:* Capabilities can be rapidly and elastically Provisioned, to quickly scale out and rapidly released to quickly scale in. To the end user, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.

- *Measured Service:* cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the Provider and consumer of the utilized service.

Cloud computing service model

As shown in figure 1 the stack model for CC consist of three major services each of them have certain responsibilities which are shifted to the cloud service provider to allowing the consumer of cloud services to focus more on their own business requirement and less on the underlying technologies.
These major services can be summarized as follow:

- Software as a Service (SaaS)

The capability provided to the customer is to use the Provider's applications running on a cloud infrastructure. the applications are accessible from various client devices through a thin client interface such as a web browser (e.g. Web-based email). The customer does not manage or control The underlying cloud infrastructure, including network, servers, operating systems, storage, or even Individual Application capabilities [2].

- Platform as a Service (PaaS)

The capability provided to the customer is to deploy onto The cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure Including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

- Infrastructure as a Service (IaaS)

The capability provided to the customer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating Systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications.

Cloud computing roles
Roles define the responsibilities, access and profile of different users that are part of a cloud computing solution. Figure 2 presents these roles defined in the three service layers [4].

The provider is responsible for managing, monitoring and guaranteeing the availability of the entire structure of the cloud computing solution. It frees the developer and the final user from such responsibilities while providing services in the three layers of the architecture, the developers use the resources provided by IaaS and PaaS to provide software services for final users.

Cloud computing deployment model

According to the intended access methods and availability of cloud computing environments, there are different models of deployment; they include private cloud, public cloud, community cloud and hybrid cloud, which are briefly analyzed in table 1.

-*Private cloud*: The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.

-*Public cloud:* the cloud infrastructure is made available to the general public and is owned by cloud service provider.

- *Community cloud:* the cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns. It may be managed by the organizations or a third party and may exist on premise or off premise.

- *Hybrid cloud*: the cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology [5].

-Green cloud computing and environmental sustainability

Green cloud computing (GCC) is the practice of implementing
policies and procedures That improve the efficiency of computing resources in such a way as to reduce the energy consumption and environmental impact of their utilization [7][8].this means how to take the maximum benefits from CC to the service providers and customers while minimize the negative impact on environment such as energy consumption and carbon emissions footprint , in order to achieve this goal we need an efficient solution to Achieve this difficult equation .

Environmental and energy conservation issues have taken center stage in the global business arena in recent years. The reality of rising energy costs and their impact on international affairs coupled with the increased concern over the global warming climate crisis and other environmental issues have shifted the social and economic consciousness of the business community.

As shown in figure3the growing demand for cloud computing Increases the energy consumption in data centers, which has become a major concern for both industry and society [9].this increase in energy consumption not only increases energy cost but also increases carbon emission. High energy-Cost results in reducing cloud providers' profit margin and high carbon emission which has negative impact on the environment [10]. Therefore we can use the Load balancing technique as energy saving solution to achieve the GCC Concept.

II. Background and related work

This section presents the various work and studies that has been done so far by the scientific community dealing with energy efficient and load balancing techniques in cloud data center.

- Energy efficient techniques

The first study discusses energy efficient computing and stress about the importance of saving energy [12]. They present the environment protection agency (EPA) report in their paper which highlights some of the following important points:

- 61 billion kWh (kilowatt hours) was consumed by servers and data center alone in 2006.
- IT equipment which is necessary to run the data center alone consumed a considerable amount of energy.

The second study try to extend the tickles Idle by introduce an Eco-friendly daemon which uses dynamic voltage and scaling to reduce the power consumption and strictly maintaining performance However, we need to predict the future workload to adjust the frequency of the CPU accordingly. If the windows size is too large, then we have to wait for a while to get the required windows size data to make an accurate prediction. Finding the right parameters to sample the workload to estimate the future workload is a challenge [13].

The third study proposed an approach for power conservation in server systems based on fast transitions between active and low power states. The goal is to minimize power consumption by a server while it is in an idle state [14]. Instead of addressing the problem of achieving energy proportional computing as proposed by barroso and holzle, the authors require only two power states (sleep and fully active) for each system component. The other requirements are fast transitions between the power states and very low power consumption in the sleep mode [15].

Load balancing techniques

The motivation of the survey of existing load balancing techniques in cloud computing is to encourage the amateur researcher to contribute in developing more efficient load balancing algorithms, this will benefit interested researchers to carry out further work .

The first study proposed a new content aware load balancing policy named as workload and client aware Policy (WCAP) it uses a unique and special property (USP) to specify the unique and special property of the requests as Well as computing nodes. USP helps the scheduler to Decide the best suitable node for the Processing the request, this strategy is implemented in a decentralized manner with Low overhead [16].

The second study proposed a join idle queue load balancing algorithm for dynamically scalable web services, this algorithm provides large scale load balancing with distributed dispatchers by, first load balancing idle processors across dispatchers for the availability of idle processors at each dispatchers and then, assigning jobs to processors to reduce average queue length at each processor, by removing the load balancing work from the critical path of request processing , it effectively reduces the system load ,incurs no communication overhead at job arrivals and does not increase actual response time [17] .

The third study proposed a new server based load balancing policy for web servers which are distributed all over the world. It helps in reducing the service response times by using a protocol that limits the reduction of Requests to the closest remote servers without overloading them; a middleware is described to implement this protocol [18].

III. Green Cloud Load Balancing Model

Based on server utilization, there are often periods of idleness where the data center servers are not used and during this period these servers still consume about 65-70% of the power consumed when it is fully utilized [19] [20] [21] [22] without doing any useful work, therefore we introduce in this paper a GCLB model to eliminate the wasted power consumption during the idle period while maintaining the QOS.

The straight forward scheme is to set the server to sleep once the queue is empty and is waken up upon a new job arrival [23] this approach works but with some limitations and it is not efficient due to the following concerns:

- The fast transition from the running mode into the sleep mode might be too costly. If the idle duration or non-idle duration is short, the server is wasting time in transition without taking enough duration of sleep (or execution of jobs).

- The average response time threshold would be always violated if the wake-up transition time is above the average Response time threshold. All jobs after the wakeup will Experience at least the wake-up transition delay; this is another big drawback of this simple approach.

In our design of energy saving schemes, we propose the following constant parameters of time periods to overcome the above raised concern:
- Time to sleep threshold (TTS) it is the minimum length of the idle duration before the server is put into the sleep power mode.
- Time to wake threshold (TTW) it is the maximum length of the sleep duration which the server can stay in the sleep power mode continuously.

-Time to delay threshold (TTD) if the job arrives earlier before the expiration of (TTW) it will remain in queue till the Expiration of (TTW) period.

-Green Cloud Load Balancing Model

Our GCLB model can be described as the following steps:
I- once the queue is empty; the server intends to stay in the idle power mode for TTS threshold.
1- If a new job arrives before the expiration of the TTS threshold, the server will immediately serve the new job. Otherwise, the server will enter to the sleep power mode and then it will be enforced to stay in the sleep power mode for TTW threshold.
2- If a new job arrives before the expiration of the TTW threshold the new job will wait in queue for TTD threshold. Otherwise, the server will continue in the sleep mode until the expiration of TTW period.
3- once the server is in the running power mode, the server runs at a constant speed ratio (R) in the running power mode serving jobs in the queue until the queue become Empty .
4- If there are servers with low utilization, their load will transferred to the medium load servers.
5- If there are servers with high utilization, the system will wake up one of the sleeping server and transfer the extra load to it.
6- Once the queue is empty, repeat Step I

The green cloud load balancing algorithm has two basic sections:
the energy section is responsible for achieving energy saving by switching idle compute server to sleep power mode, and if the compute server is using less than30% of its resources, GCLB Algorithm transfer their load to one of The medium load server. It does this by first gathering the utilization percentage of each active compute server.

The balancing section is responsible for maintaining QOS and response time for job requests by wake up one of sleeping nodes that can help the overloaded servers to lighten its load, and this is happened if one of the servers has utilization over (65%).
the threshold of (65%) utilization was chosen since when (35%) of the resources are available, at least one more virtual machine can be accommodated using three out of five available configurations.

IV-Research Methodology

This section presents the general view of the system architecture followed by a more detailed view of the component's architecture additionally; it shows the measurement's system that describes the power, Performance and CPU utilization's measurements. Then, the experimental methodology will be presented.
in order to achieve the objectives of this research and due to the nature of the study we opt to employ cloud Based E-learning model as shown in Figure 5, where the new system model is being constructed and evaluated, we find it useful to employ the modeling and experimental methodology, because such methodology enables us to describe the proposed system model and further validate it through experiment processes where it becomes possible to compare and analysis results between the current and proposed model.
As shown in figure 6, an actual cloud system e-learning model has been implemented we find it is useful to use the measurement and experimenting methodology for collecting, analysis and validating the proposed GCLB model.

- **Data gathering methods**

The overall approach to this thesis is to collect useful, quantifiable data to conclude whether or not increasing the server utilization efficiency and sleeping idle servers will be advantageous in reducing power consumption and maintain the response time.

The following areas are explored to gather sufficient data for a power saving analysis.

-CPU Utilization (Usage)

CPU utilization is a key performance metric and major indicator of power consumption. it can be used to track CPU performance regressions or improvements, and is a useful data point for performance problem investigations.

In this paper the proposed GCLB algorithm will depend on an idea of setting a CPU utilization threshold to distinguishing between the under-Loaded and over-loaded states of the host.
When the algorithm is invoked, it compares the current CPU utilization of the host with the defined threshold. If the threshold is exceeded the algorithm detects a host overloaded and vice versa.

We will use Processor :(%) Processor Time counter in windows performance monitor built-in tool to monitor the amount of time the CPU spends processing a non-idle thread.

- Power consumption

We will use the power metering and budgeting (PMB) tool to measure and monitor the power consumption in order to evaluate and validate the proposed GCLB model.
a power meter is a hardware component of the system that reports information about power consumption, in watts. This information is typically provided as part of a power supply or by using a base board management controller.
This infrastructure tool promotes energy efficiency on cloud data center servers by providing power consumption and management features, additionally PMB provides extra options to configure power metering and budgeting to help cloud provider and end users to balance power and performance to meet their needs.

The PMB infrastructure feature provides two type of information:

Power Metering Information

This information is used to determine how the computer system or sub-components use power. Power consumption is monitored, or metered, by a power meter in the system. Power metering also provides the current configuration of a power meter, such as metering capabilities and power consumption thresholds.

Power Budgeting Information

This information is used to determine the power limit, or budget, that is supported by the computer system. Depending on the hardware platform, this information might also allow you to configure the system's power budget.

Performance (Throughput)

Throughput is a measure of how many units of information a system can process in a given amount of time. We will use System\% user time and System\% privileged time counter in windows performance monitor tool to measure the percentage of time the processor is spending executing User processes.

- Data analysis Methods

Due to the nature of the quantitative research type and the numerical data which collected through the above data collection methods, we find it is useful to use the graphical chart and comparison table methods to analysis and validate the collected data.
We will compare and analysis the Power consumption, CPU utilization and throughout data which collected before and After we implement the proposed GCLB model and display The results through graphical chart and comparison table to discover and understand the relationship between the number of users, CPU utilization and power consumption and how the proposed model can achieve energy efficiency while maintain the performance and server availability.

- Experimental Setup

In this section, we will discuss the cloud E-learning model which will use in our experimental setup to evaluate and validate the proposed GCLB Model, we discuss the software and hardware used and the modifications made to them as shown in Figure 6.

-**Hardware required**

-Cloud Host server (Hyper V)

The study will use dell precision T5400 blade server, with processor Intel (R) Xeon (R) CPU X5450@ 3.00GHz 2.99GHz (4 processors), installed memory 16 GB and system type: 64-bit operating system, x64-based processor.

-Web Front End Servers (Guest Nodes)

The study will use three dell precision T5400 workstation, with processor Intel (R) Xeon (R) CPU X5450@ 3.00GHz 2.99GHz (2 processors), installed memory 6 GB and system Type: 64-bit operating system, x64-based processor.

-**Software required**

-Cloud Host server (Hyper V)

We will use windows server 2012 R2 data center with Microsoft Hyper V 2012 feature.
Web Front End Servers (Guest Nodes)

We will install on each web front end server, windows server 2008 R2, Microsoft Network load balancing (NLB) 2012, PHP 5.3 None thread safe installer, PHP Manager 1.2 for IIS 7.5, SQL server 2012 and Moodle 2.6.

V- Experimental results and discussion

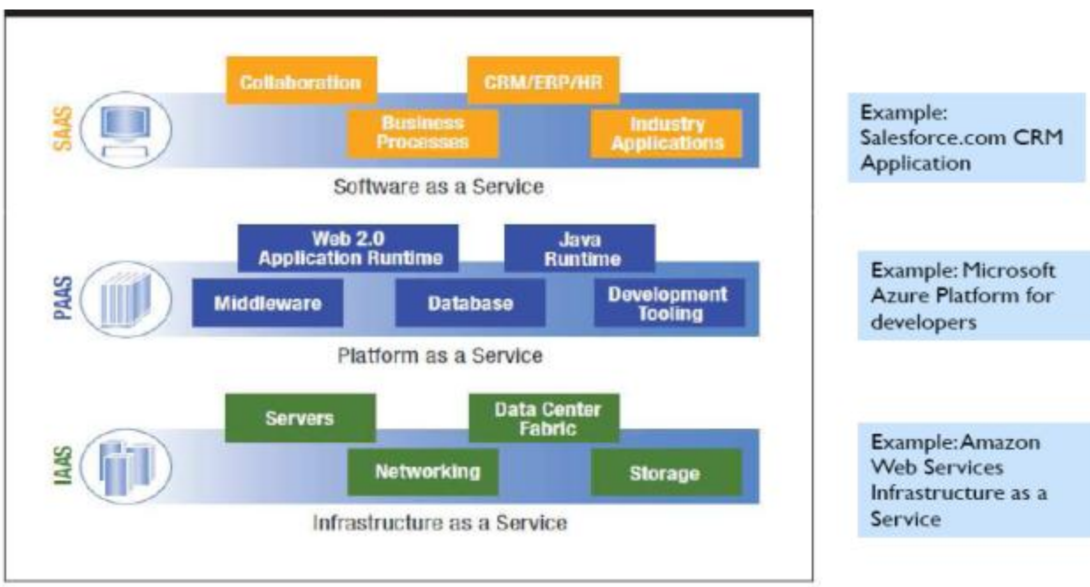

Fig. 1: Cloud computing stack model
(Source: Mather, Kumaraswamy and Latif, 2009)

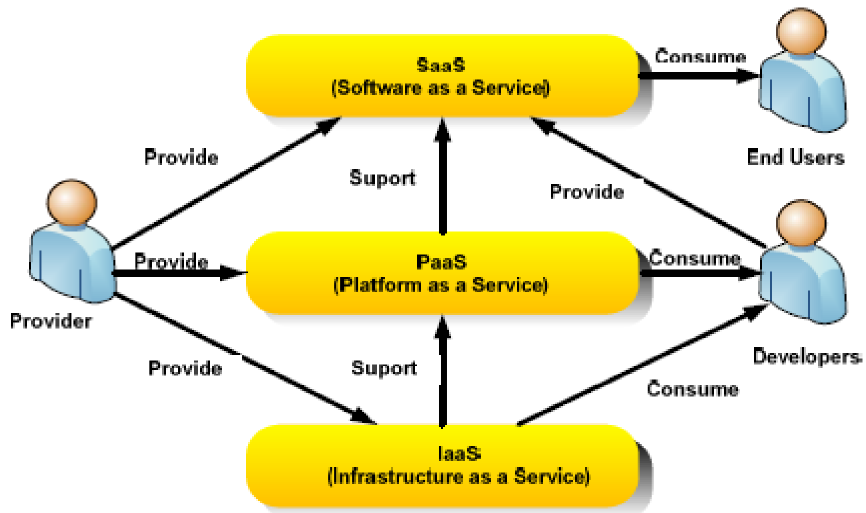

Fig. 2: Cloud computing roles
(Source: Marinos and Briscoe, 2009)

	Infrastructure Managed By[1]	Infrastructure Owned By[2]	Infrastructure Located[3]	Accessible and Consumed By[4]
Public	Third Party Provider	Third Party Provider	Off-Premise	Untrusted
Private/ Community	Organization Or Third Party Provider	Organization / Third Party Provider	On-Premise / Off-Premise	Trusted
Hybrid	Both Organization & Third Party Provider	Both Organization & Third Party Provider	Both On-Premise & Off-Premise	Trusted & Untrusted

Table.1: Cloud Computing Deployment Models.

(Source: Cloud security alliance team, 2009)

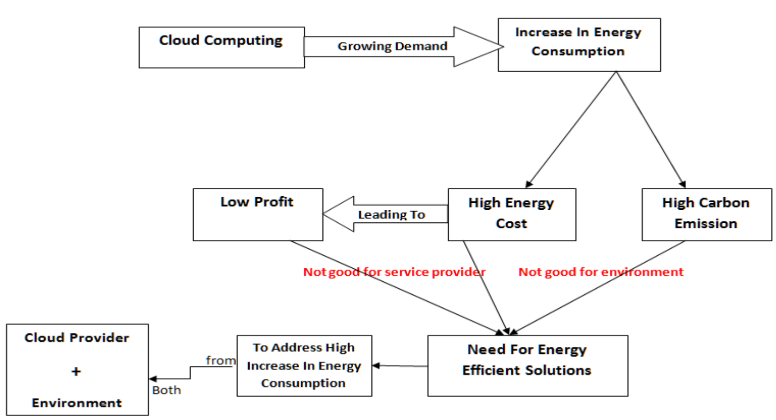

Fig. 3: Cloud and environmental sustainability
(Source: Kansal and Chana, 2012)

Energy Saving
1-for all compute servers S do
2-calculate current CPU utilization for each compute server S
3-end for
4-if Utilization=0% // this compute node is idle
5-if Tidle ≥ TTS
6-put the compute server into sleep mode
7-else
8- if Utilization ≤30% // this compute server is under- loaded
9-transfer the load to the medium –loaded servers
10-end for
balance
11-if Utilization >65% utilization // this compute server is over- loaded
12-Wake up one of the sleeping compute node
13-Then
14-Transfer extra load from over-loaded server to new wake-up server
15-end if

Fig.4: Green cloud load balancing algorthim

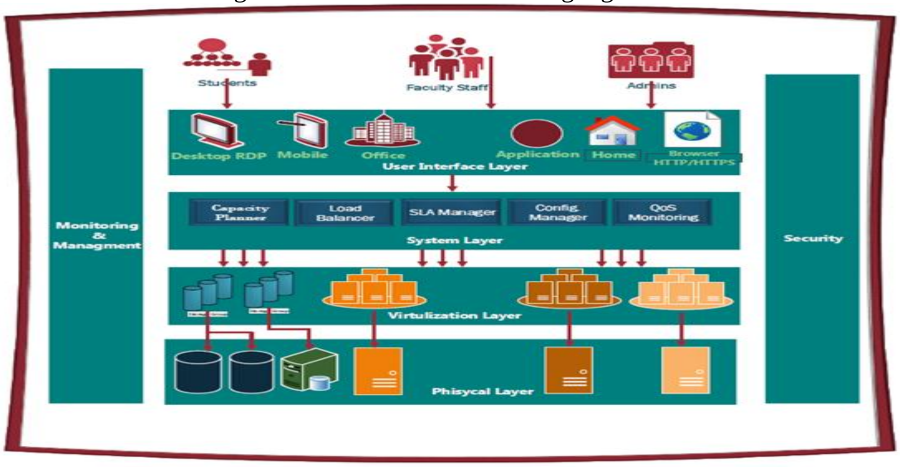

Fig.5: Cloud based E-learning model

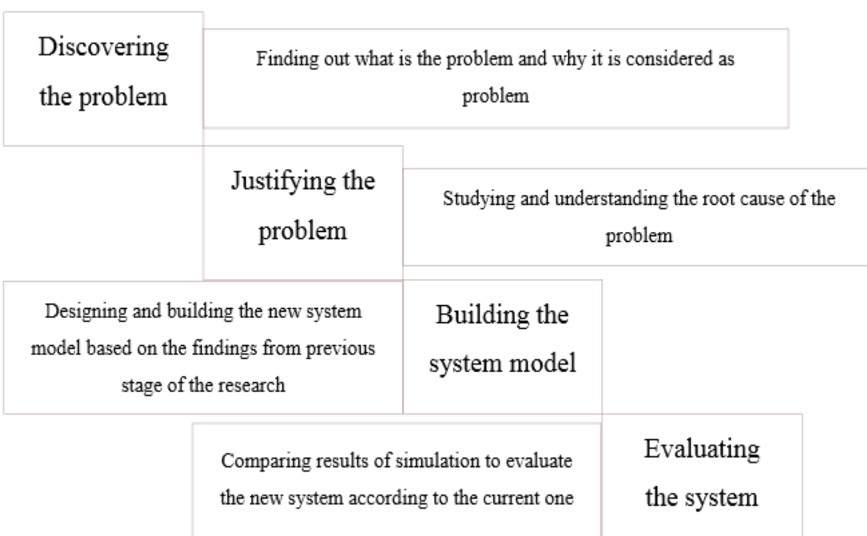

Fig .6: Research methodology steps

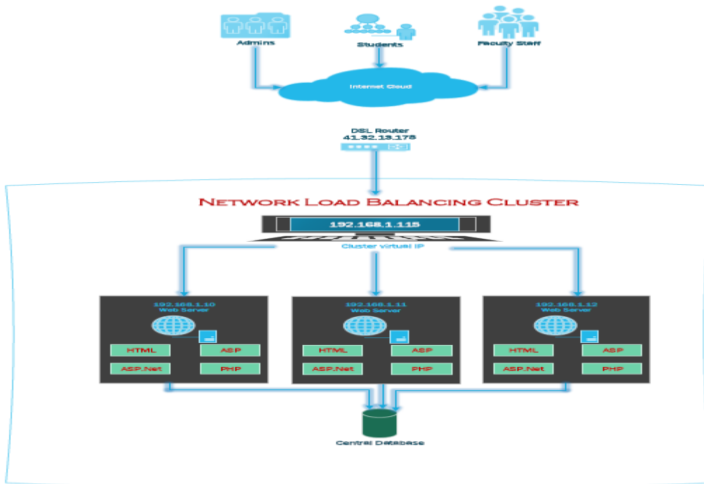

Fig 7: Experimental setup Model

This part presents the results of the experiments that were conducted based on the methods described in pervious section, The data obtained throughout the experiments were analyzed and interpreted .summaries of results are generally presented in figures. Typical graphical charts relating to the experiments will be provided in this paper to analyze the relationship between the energy consumption and values of the utilization threshold. The values of the result show that the lower threshold has higher influence on the energy consumption than the upper threshold. This can be explained by the fact that an increase of the lower threshold by implementing the proposed GCLB eliminates the low utilization of the resources leading to higher energy saving.

For the benchmark experimental results we have used a Normal Load Balancing (NLB) policy. This policy does not apply any power saving optimizations and implies that Workload will distribute equally between all nodes regardless the CPU utilization percentage and the server idle periods. The second policy applies GCLB model which eliminating the servers idle period by sleeping the data center servers during idle periods based on the CPU utilization and GCLB algorithm.
Comparative results of the energy consumption and average throughout times for each of the two policies with different sets of request arrival rates

The NLB policy exhibit higher energy and power consumption, the reason is primarily due to the fact that servers once activated, stay in operational or active state throughout the duration of the run. The idle power of activated servers when not in use, contributes to the substantially high power and energy consumption.
Using our power saving based approach, energy and power consumption of the cluster is reduced and gives a savings of at least 40% when compared with NLB baseline policy during idle period, as shown in figure 15 and 18

On response times, Normal Load Balancing (NLB) Baseline policy perform the best because, servers are more often in operational state for the requests to be processed compared to GCLB schemes .

-Comparison of GCLB and Normal Load Balancer NLB

The results obtained by using our GCLB algorithm would obviously bring higher power savings over conventional load balancing algorithm, In the normal load balancing approach, the cluster controller assigns virtual machine requests to compute nodes sequentially. This effectively balances the load across the compute nodes, but leaves these nodes in an “on” state with usually low utilization. Figure 1 shows the power consumption of the Normal load balancing algorithm.
The major performance difference between our GCLB algorithm and the Normal Load balancing approach is that compute nodes that are idle using GCLB are switched to the low power sleep mode. The Normal Load balancing approach always keeps all compute nodes powered on, no matter the number of active virtual machines. The Normal Load balancing approach is effective in load balancing across the available compute nodes, but such a relatively simple load balancer consumes a large amount of unnecessary power.

We will use our experimental cloud Moodle to monitor the power consumption, response time, disk and memory for web front end servers under different workload before and after applying the Green Cloud Load Balancing (GCLB) model.

Finding analysis and discussion

-power consumption saving during peak period

After we measure power consumption for cloud host server and during *peak* period before and after we implement the GCLB model as shown in figure 11 and 12 the power consumption reduced from 155 watt to 149 watt with power Saving percentage (4 %)

- Power consumption saving during idle period

After we measure power consumption for cloud host server during idle period before and after we implement the GCLB Model as shown in figure 15 and 16, the power consumption reduced from 167 watt to 100 watt with power saving Percentage (40 %).

The GCLB algorithm is more effective in reducing power consumption during idle period due to the low CPU Utilization and huge power wasted which can be eliminated through put the idle servers in sleep power mode state and consolidate the workload on the active servers which Increase CPU usage and reduce power consumption.

- Throughput

After we measure the throughput for cloud server before and after we implement the GCLB model during peak period as shown in figure 10 and 11 and during idle period as shown in figure 15 and 16 the proposed GCLB model Maintain the throughput and the cloud performance without SLA violation.

GCLB VS NLB during peak period

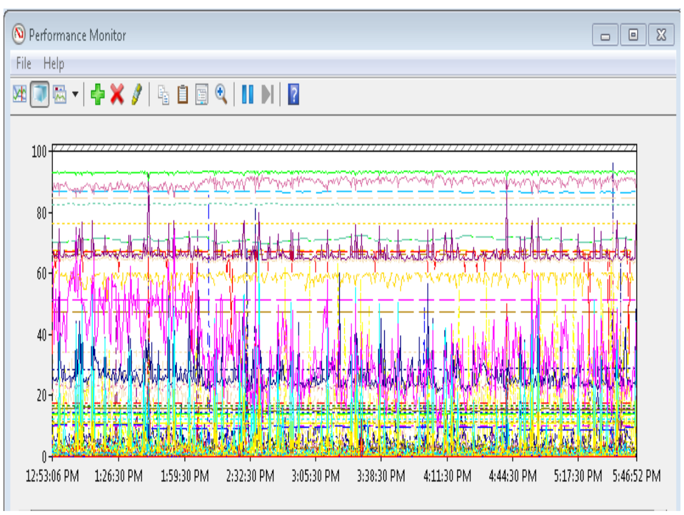

Fig .8: Cloud Server during peak period before GCLB

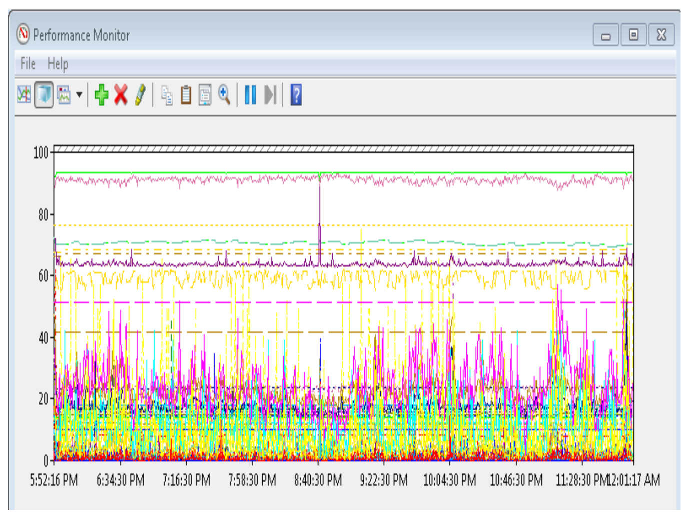

Fig 9: cloud server during peak period after GCLB

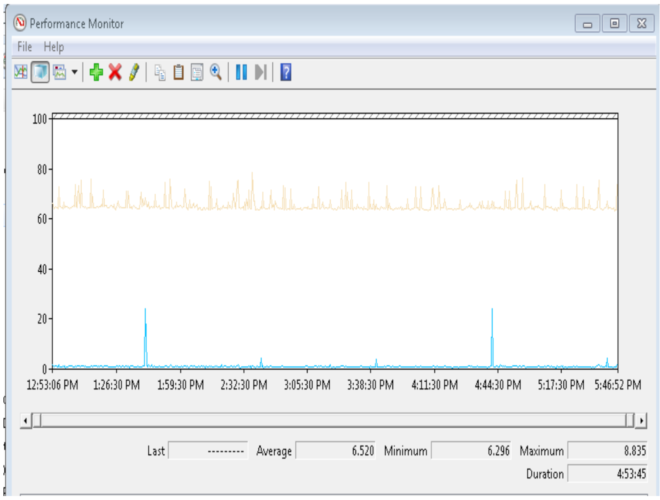

Fig 10: Throughput during peak period before GCLB

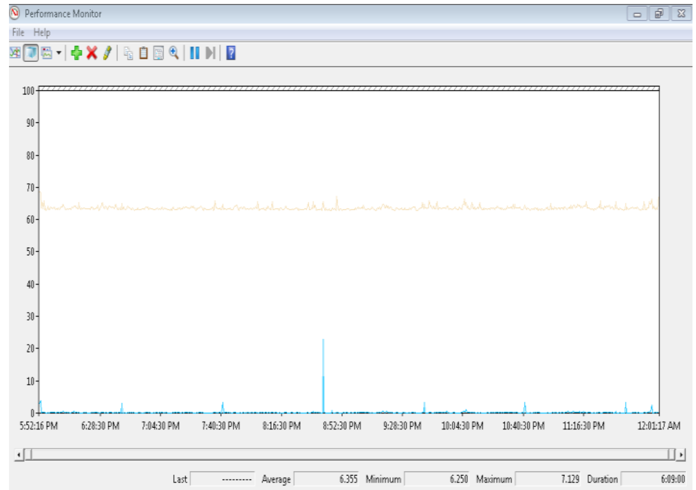

Fig 11: Throughput during peak period after GCLB

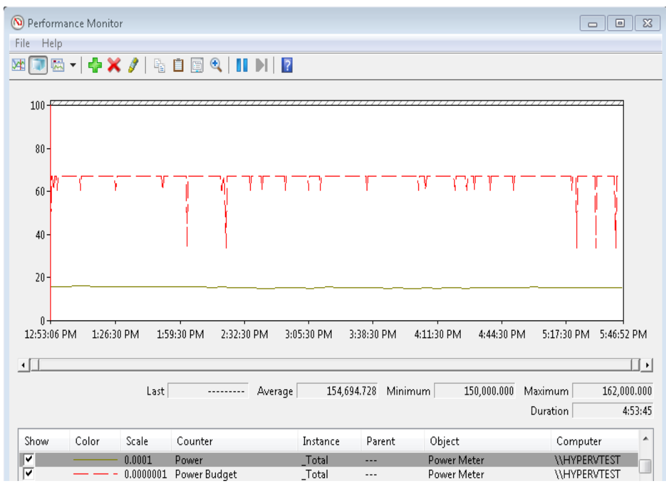

Fig.12: Power consumption during peak period before GCLB

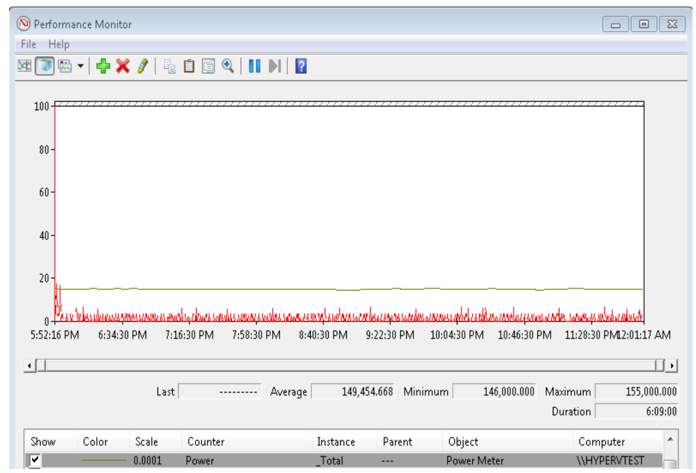

Fig 13: Power consumption during peak period after GCLB

GCLB VS NLB during idle period

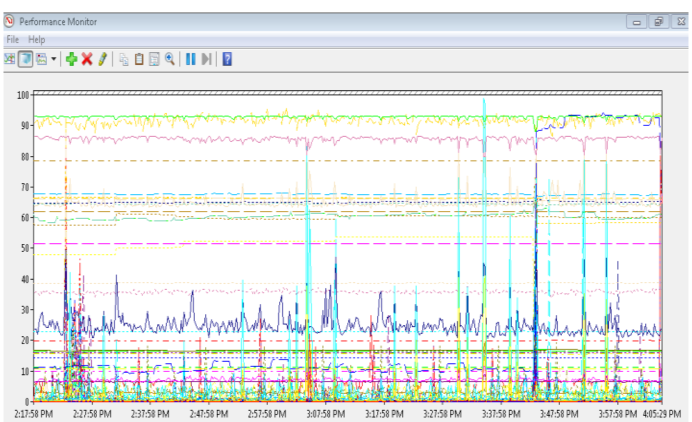

Fig 14: Cloud Server during idle period before GCLB

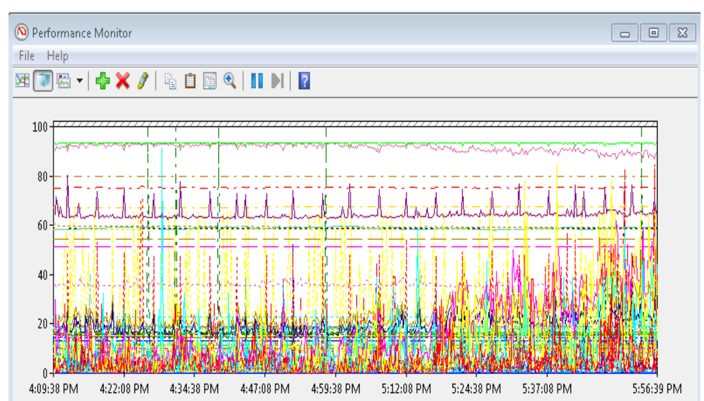

Fig 15: Cloud Server during idle period after GCLB

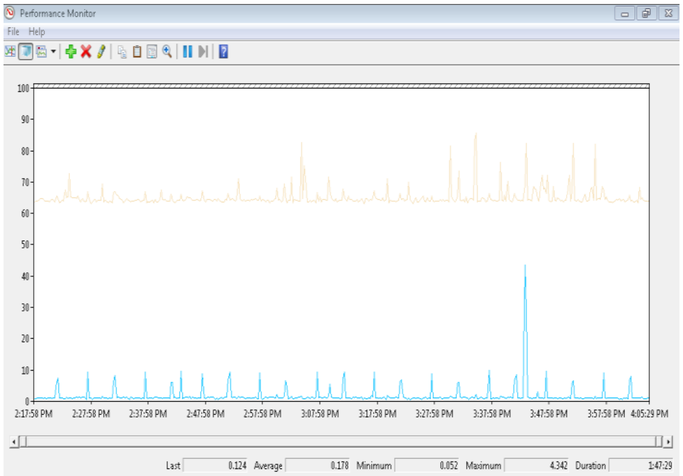

Fig 16: Throughput during idle period before GCLB

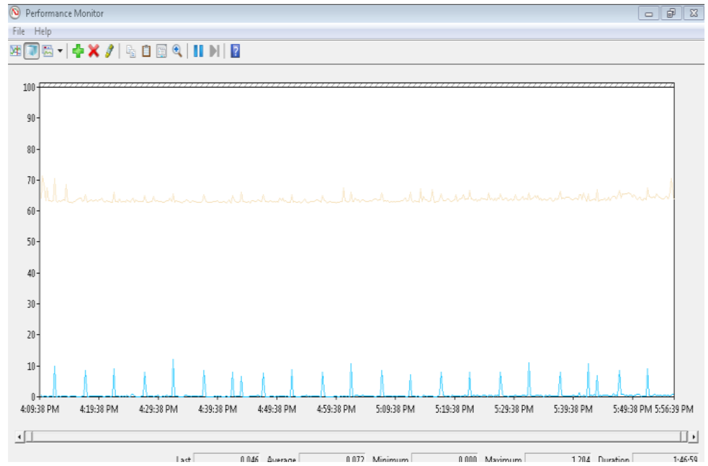

Fig 17: Throughput during idle period after GCLB

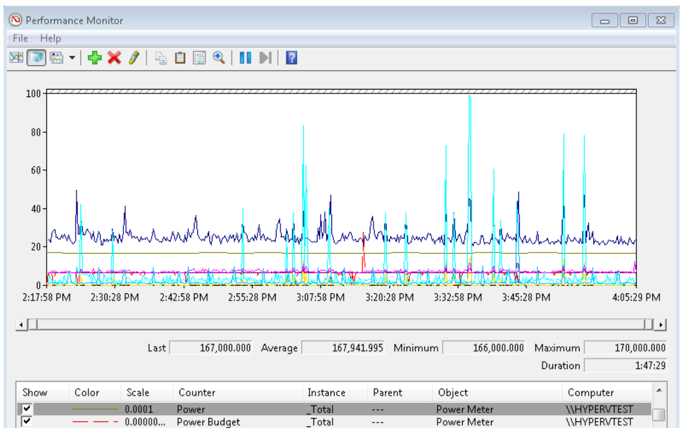

Fig 18: Power consumption during idle period before GCLB

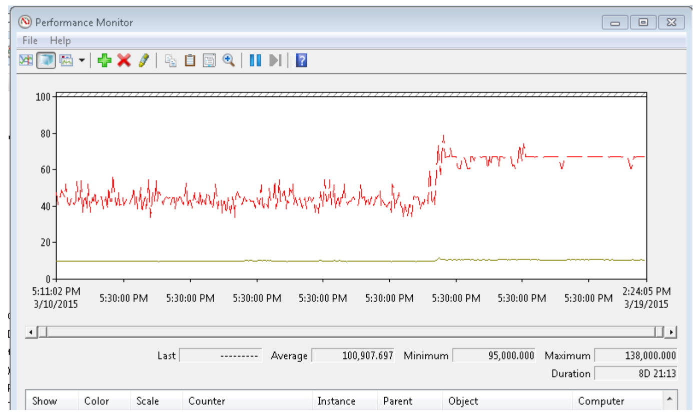

Fig 19: Power consumption during idle period after GCLB

Comparative analysis before and after GCLB

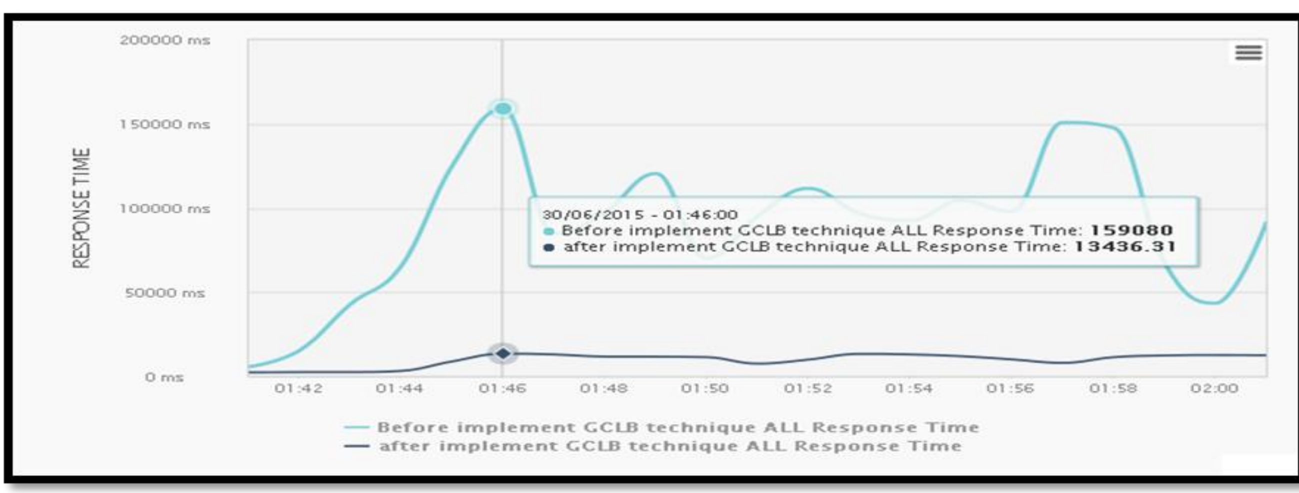

Fig 20: Response time before and after GCLB

As shown in figure 20, the amount of time system taken to process a request after it has received one after implementing GCLB is better than the amount of time the system takes before implementing GCLB. This is due to high disk, memory and CPU utilization for the cloud data center after applying the GCLB model.

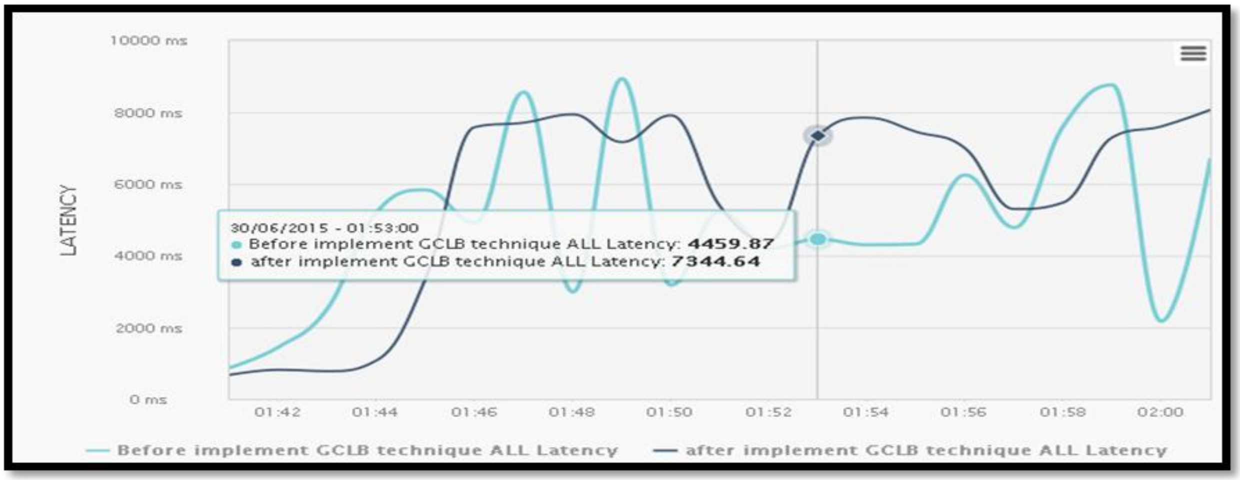

Fig 21: Latency before and after GCLB

The figure 21 shows that the delay involved for your request to reach to server after implementing GCLB is lesser than the delay involved for your request before implementing GCLB. This is due to the new load balancing technique which distributes the incoming user's requests among the cloud data center servers that reduce the delay and enhance the response time.

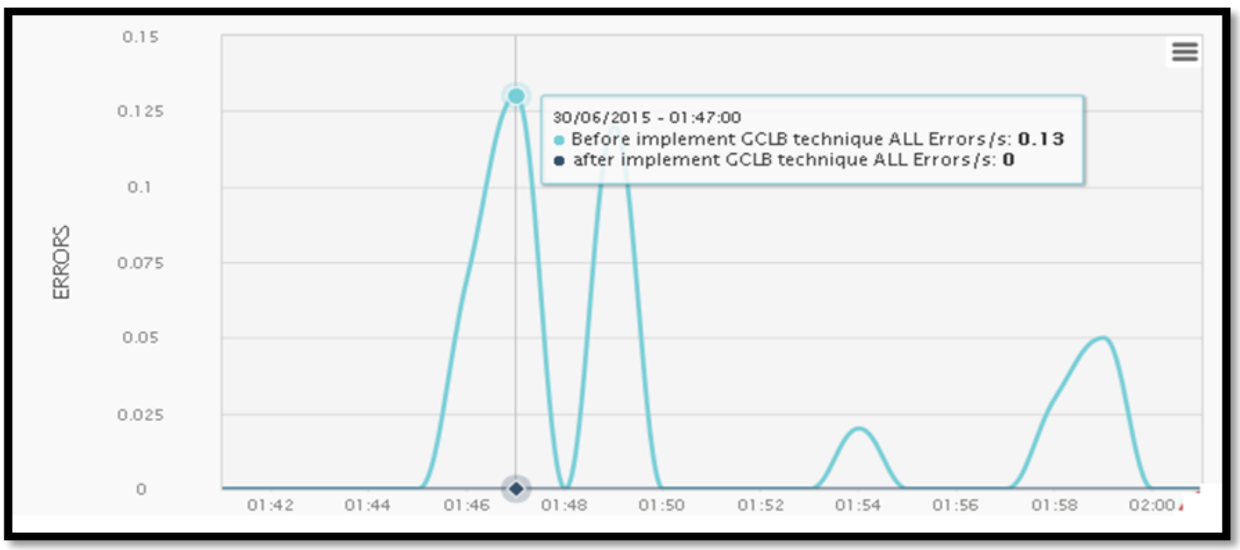

Fig 22: Error before and after GCLB

Figure 22 shows that zero error after implementing GCLB model is rather than the huge error before implementing GCLB model which affects the quality of service and causes a delay in cloud server response.

References

1- P. Mell and T. Grance. “The NIST definition of cloud computing”., pp.1-106, (2011) .
2- M .AL-Roomi, S. Al -Ebrahim, S. rais and I. Ahmad, “Cloud Computing Pricing Models: A Survey” International Journal of Grid and Distributed Computing Vol.6, No.5 (2013), pp.93-106.
3- T. Mather,S. Kumaraswamy, S. Latif “Cloud Security and privacy an Enterprise perspective on Risks and Compliance”, 2009.
4- A.Marinos and G. Briscoe, “Community cloud computing,” in First International Conference Cloud Computing, CloudCom, volume 5931 of Lecture Notes in Computer Science, pp. 472–484. Springer (2009).
5- C. Rao,M. Leelarani and Y. Kumar“cloud computing services and deployment models”, International Journal Of Engineering And Computer Science, volume 2 issue 12, pp.3389-3392 ,December 2013 .
6- *Cloud security alliance team,* “*Security guidance for critical areas of focus in cloud computing* ”,available *From:https://cloudsecurityalliance.org/guidance/csaguide.v*1.0.pdf [accessed April.2009]
7- S. Kabiraj, V. Topka, and R. C. Walke, “Going Green: A Holistic Approach to Transform Business”, International Journal of Managing Information Technology (IJMIT), Vol. 2, No. 3, August 2010, pp. 22-31.
8- J. Baliga, R. W. A. Ayre, K. Hinton, and R. S. Tucker,“Green Cloud Computing: Balancing Energy in Processing, Storage, and Transport”, Proceedings of the IEEE, Vol. 99, No. 1, January 2011, pp. 149-167.
9- K. M. Nagothu, B. Kelley, J. Prevost, and M. Jamshidi, “Ultra low energy cloud computing using adaptive load prediction”, Proceedings of IEEE World Automation Congress(WAC) , Kobe, September 2010, pp. 1-7.
10- S. K. Garg, C. S. Yeob, A. Anandasivamc, and R. Buyya, “Environment-conscious scheduling of HPC applications on distributed Cloud-oriented data centers”, Journal of Parallel and Distributed Computing, Elsevier, Vol. 70, No. 6, May 2010, pp. 1-18.
11- N. Kansal and I. Chana “Cloud Load Balancing Techniques: a step towards green Computing”, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 1, January 2012 .
12- R. Brown, E. Masanet, B. Nordman and B. Tschudi,, “Report to Congress on Server and Data Center Energy Efficiency“, Public Law, July 2007,pp.109-431.
13- S .Huang, W. Feng, “A workload-aware, eco-friendly daemon for cluster computing“, Technical Report, Computer Science, Virginia Tech (2008).
14- D. Meisner, B. T. Gold, and T. F. Wenisch, Power Nap: eliminating server idle power, ACM SIGPLAN Notices, vol. 44, no. 3, pp. 205–216, 2009.
15- L. Barroso and U. Holzle, “The case for energy-proportional computing,” Computer, vol. 40, pp. 33 – 37, Dec 2007.
16- H. Mehta, P. Kanungo, and M. Chandwani, “Decentralized content aware load balancing algorithm for distributed computing environments”, Proceedings of the International Conference Workshop on Emerging Trends in Technology (ICWET), February 2011, pp. 370-375.
17- Y. Xiea Q. Kliotb G. Gellerb A.Larusb and A.Greenber, “Comparative Analysis of Existing Dynamic Load Balancing Techniques”, International Journal of Computer Applications, IJCA 70(26): pp. 25-29, May 2013.
18- Nakai A.M., Madeira E. and Buzato L.E. (2011) 5th Latin-American Symposium on Dependable Computing, PP.156-165
19- A .Greenberg, J .Hamilton, D .Maltz,and P .Patel, “The cost of a cloud: research problems in data center Networks”, ACM SIGCOMM Computer communication Review, Volume 39 Issue 1, January 2009, pp. 68- 73
20- C. Lefurgy, X. Wang, and M. Ware, “Server-level power control”, ICAC’07: International Conference on Autonomic Computing, June 2007.
21- X. Fan, W. Weber, and L. Barroso, “Power provisioning for a warehouse-sized computer”. ISCA ’07: International Symposium on Computer Architecture, 2007.
22 -L. Barroso and U. Holzle, “The Case for Energy- Proportional Computing”, IEEE Computer, pp. 33–37, December 2007.
23- A.Gandhi, V.Gupta, M.Harchol-Balter, and M.Kozuch, “Optimality analysis of energy-performance trade-off for server farm management”, Performance Evaluation, vol. 67 Issue 11, pp. 1155-1171, November, 2010.