



RESOURCE OPTIMIZATION STRATEGY FOR PERFORMANCE IMPROVEMENT IN VIRTUALIZED CLOUD ENVIRONMENT

S.Kalaiselvi , C.S.Kanimozhi Selvi ,R.S.Mohana
Assistant Professor(SLG), Associate Professor, Assistant Professor
Department of CSE, Kongu Engineering College, Erode, Tamilnadu

kalaiselvisubbarayan@gmail.com, kanimozhi@kongu.ac.in, mohanapragash@kongu.ac.in

Abstract— Cloud computing is a cost-effective model for provisioning services and makes the ICT management effective in dealing with real-world workloads. Most of the organization can either outsource its computational needs to the cloud avoiding high up-front investments in a private computing infrastructure and consequent maintenance costs, or implement a private cloud data center to improve the resource management and provisioning processes. The global energy consumption of IT equipment is steadily rising and produces an increasing portion of global energy production. Currently, data centers consume about 1.5% of global electricity production, whereby their total energy usage has almost tripled between 2000 and 2010. The increasing demand of computational power, especially in current cloud computing environments, is an important reason for the rising number of running computers as well as increasing energy consumption.

Index Terms—Cloud Computing, Energy Consumption, Load Balancing.

I. INTRODUCTION

In recent year, cloud computing has become a popular computing paradigm for hosting and delivering services over the Internet. The adoption and deployment of cloud computing platforms have many attractive benefits, such as reliability, quality of service and robustness. To the consumer, the cloud appears to be infinite, and the consumer can purchase as much or as little computing power as they need. From a provider's perspective, the key issue is to maximize profits by minimizing the operational costs. In this regard, power management in cloud data centers is becoming a crucial issue since it dominates the operational costs. Moreover, power consumption in large-scale computer systems like clouds also raises many other serious issues including the emission of carbon dioxide and system reliability.

II. ENERGY CONSUMPTION IN DATA CENTERS

How to reduce power consumption of large scale cloud data centers has received worldwide attention. Reducing energy consumption is a critical step in

lowering data center operating costs for various data centers. As such, with the growing popularity of cloud computing, it is necessary to examine various methods by which energy consumption in cloud environments can be reduced.

Although, the Data centers adopt the best practices in design, operation and maintenance to achieve operational excellence, they have to adopt some technologies like virtualization to save energy. Now, the data centers are looking for solutions that are more intelligent, in order to achieve energy efficiency and manage the resources. Virtualization in Data center offers various benefits like increased resource utilization, decreased power and cooling consumption, faster provisioning, and saving in space requirements. The proposal covers the energy saving opportunities through adoption of virtualization of servers & network component.

Nowadays Cloud Computing is utilized in almost all fields like Medical, Education, Entertainment, business etc. In particular, cloud computing-based solutions should be considered for applications that require the agility, scale-out, and ability to integrate and analyze massive data that cloud computing can provide. Examples of such applications include: big data analysis and all-source intelligence integration; processing, exploitation, and dissemination of data gathered through intelligence, surveillance and reconnaissance; large scale modeling and simulation; open source data collection, storage, and assessment; and advanced decision support systems.

As reported by the Open Compute Project, Facebook's Oregon data center achieves a Power Usage Effectiveness of 1.08, which means that 93 percent of the data center's energy consumption are consumed by the computing resources. Therefore, now it is important to focus on the resource management



aspect i.e., ensuring that the computing resources are efficiently utilized to serve applications so that energy can be consumed optimally. Recently, the problem of energy consumption by data centers is addressed by many researchers. Xiaoying Wang, Zhihui Du and Yinong Chen[1] investigated adaptive model free approaches for resource allocation and energy management under time varying workloads and heterogeneous multi-tier applications. The objective of their research is to guarantee the summarized revenue of the resource provider while saving energy and operational cost.

Zhen xiao, Weijia Song and Qi Chen[3] suggested an approach that uses virtualization technology to allocate data center resources based on the application demands and support green computing by optimizing the number of servers in use. Also, they developed a set of heuristics that prevent overload in the system effectively while saving energy.

It is well known that the scope for energy conservation in India is enormous. There are huge gains to be made by upgrading technology, equipment and appliances in a wide range of sectors. The service sector has been experiencing a significant growth in India and major part of this is attributable to the IT sector. High tech facilities in this sector are making it one of the fastest growing energy use sectors.

According to a survey conducted by Central Electricity Authority of India and by the report of Energy Statistics 2011(under Ministry of Statistics and Programme Implementation), the power consumption of Industry is indicated as 45.52% of total energy share. Although net consumption of data centers in India has yet to be quantified, the IT-heavy nature of the country's economy suggests relative data center consumption in India may contribute much among all sectors.

The resource utilization and QoS is explained by the fact that server overload cause resource shortage and performance degradation of applications. Anton Beloglazov and Rajkumar Buyaa[9] proposed a novel approach to manage the overloaded hosts that optimally solves the problem of post overload detection by maximizing the mean inter migration time under the specified QoS. The objective of the proposal is that minimizing energy consumption and maximizes the QoS delivered by the system which form an energy-performance tradeoff. Christo Ananth et al. [11] discussed about a method, In vehicular ad hoc networks (VANETs), because of the nonexistence of end-to-end connections, it is essential that nodes take advantage of connection opportunities to forward messages to make end-to-end messaging possible. Thus, it is crucial to make sure that nodes have incentives to forward

messages for others, despite the fact that the routing protocols in VANETs are different from traditional end-to-end routing protocols. In this paper, stimulation of message forwarding in VANETs is concerned. This approach is based on coalitional game theory, particularly, an incentive scheme for VANETs is proposed and with this scheme, following the routing protocol is in the best interest of each node. In addition, a lightweight approach is proposed for taking the limited storage space of each node into consideration.

Google, Salesforce, IBM, Microsoft, and Oracle have begun to establish new data centers for hosting cloud computing applications in various locations around the world to provide redundancy and ensure reliability in case of site failures. Due to the gain of carbon dioxide pollution in our environment, rising energy costs and future high-growth IT market, approaches are needed to improve energy efficiency of the entire IT. Some methods, e.g effective reuse of waste heat, correct dimensioning of IT components and employing energy-efficient hardware are state of the art.

As part of energy efficiency initiative in Indian data centers, Confederation of Indian Industry (CII) has published energy efficiency guidelines & Best practices to be adopted in Indian data Centers which was supported by Bureau of Energy Efficiency (BEE) Ministry of Power, Government of India. Among several energy saving technologies, the adoption of effective server virtualization techniques would improve the power saving potential of the data centers by 40%.

A data center with a PUE of 2.0 ,only 50% of the power can be used on servers. Therefore, it becomes critical whether servers have used all the energy to complete the workload. We are aware that low energy utilization of a server is mainly due to its idle state caused by low CPU utilization. Even at a very low load, such as 10% CPU utilization, the power consumed is over 50% of the peak power. Thus, the energy efficiency of servers plays an important role for the entire energy efficiency of the data center.

With effective load prediction algorithms, we can proactively prevent server overloading situation and also it minimizes unnecessary migrations due to temporary load fluctuation. Based on the prediction algorithms, the resource optimization is done with the Genetic algorithms and the system will proceed the adjustment strategies according to the real-time demand of virtual machines. Once a new request of VM arrives, the system will run genetic algorithm to adjust the overall allocation of the resources.



III. EXISTING LOAD BALANCING TECHNIQUES IN CLOUDS

Following load balancing techniques are currently prevalent in clouds

VectorDot[2] is a novel load balancing algorithm called VectorDot. It handles the hierarchical complexity of the data-center and multidimensionality of resource loads across servers, network switches, and storage in an agile data center that has integrated server and storage virtualization technologies. VectorDot uses dot product to distinguish nodes based on the item requirements and helps in removing overloads on servers, switches and storage nodes.

CARTON- R. Stanojevic et al. [2] proposed a mechanism CARTON for cloud control that unifies the use of LB and DRL. LB (Load Balancing) is used to equally distribute the jobs to different servers so that the associated costs can be minimized and DRL (Distributed Rate Limiting) is used to make sure that the resources are distributed in a way to keep a fair resource allocation. DRL also adapts to server capacities for the dynamic workloads so that performance levels at all servers are equal. With very low computation and communication overhead, this algorithm is simple and easy to implement.

addressed the problem of intra-cloud load balancing amongst physical hosts by adaptive live migration of virtual machines. A load balancing model is designed and implemented to reduce virtual machines' migration time by shared storage, to balance load amongst servers according to their processor or IO usage, etc. and to keep virtual machines' zero-downtime in the process

Event-driven[10]-The event-driven load balancing algorithm used for real-time Massively Multiplayer Online Games (MMOG). This algorithm after receiving capacity events as input, analyzes its components in context of the resources and the global state of the game session, thereby generating the game session load balancing actions. It is capable of scaling up and down a game session on multiple resources according to the variable user load but has occasional QoS breaches.

Scheduling strategy on LB of VM resources –It is related to a scheduling strategy on load balancing of VM resources that uses historical data and current state of the system. This strategy achieves the best load balancing and reduced dynamic migration by using a genetic algorithm. It helps in resolving the issue of load imbalance and high cost of migration thus achieving better resource utilization.

Nidhi Jain Kansal and Indervereer Chana Task Scheduling based on LB- Y. Fang et al. [12] discussed a

two-level task scheduling mechanism based on load balancing to meet dynamic requirements of users and obtain high resource utilization. It achieves load balancing by first map-ping tasks to virtual machines and then virtual machines to host resources thereby improving the task response time, resource utilization and overall performance of the cloud computing environment.

Honeybee Foraging Behavior[12] provides decentralized honeybee-based load balancing technique that is a nature-inspired algorithm for self-organization. It achieves global load balancing through local server actions. Performance of the system is enhanced with increased system diversity but throughput is not increased with an increase in system size. It is best suited for the conditions where the diverse population of service types is required.

Biased Random Sampling[10]- It is a distributed and scalable load balancing approach that uses random sampling of the system domain to achieve self-organization thus balancing the load across all nodes of the system. The performance of the system is improved with high and similar population of resources thus resulting in an increased throughput by effectively utilizing the increased system resources. It is degraded with an increase in population diversity.

Active Clustering-[10] It is a self-aggregation load balancing technique that is a self-aggregation algorithm to optimize job assignments by connecting similar services using local re-wiring. The performance of the system is enhanced with high resources thereby increasing the throughput by using these resources effectively. It is degraded with an increase in system diversity.

(OLB + LBMM)- S.-C. Wang et al. proposed a two-phase scheduling algorithm that combines OLB (Opportunistic Load Balancing) and LBMM (Load Balance Min-Min) scheduling algorithms to utilize better executing efficiency and maintain the load balancing of the system. OLB scheduling algorithm, keeps every node in working state to achieve the goal of load balance and LBMM scheduling algorithm is utilized to mini-mize the execution time of each task on the node thereby minimizing the overall completion time. This combined approach hence helps in an efficient utilization of resources and enhances the work efficiency.

centralized content aware- H. Mehta et al. [12] proposed a new content aware load balancing policy named as work-load and client aware policy (WCAP). It uses a parameter named as 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 processing the requests. This strategy is implemented in a decentralized manner with low overhead. By using the content information to narrow down the search, it improves the searching performance overall performance of the system. It also helps in reducing the idle time of the computing nodes hence improving their utilization.

Server-based LB for Internet distributed services- [10] is 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 redirection of requests to the closest remote servers without overloading them. A middleware is described to implement this protocol. It also uses a heuristic to help web servers to endure overloads.

Lock-free multiprocessing solution for LB[12]- a lock-free multiprocessing load balancing solution that avoids the use of shared memory in contrast to other multiprocessing load balancing solutions which use shared memory and lock to maintain a user session. It is achieved by modifying Linux kernel. This solution helps in improving the overall performance of load balancer in a multi-core environment by running multiple load-balancing processes in one load balancer.

IV. METRICS FOR RESOURCE MANAGEMENT IN CLOUDS

Various metrics considered in existing resource management techniques in cloud computing are discussed below

Throughput is used to calculate the no. of tasks whose execution has been completed. It should be high to improve the performance of the system.

Overhead Associated determines the amount of overhead involved while implementing a load-balancing algorithm. It is composed of overhead due to movement of tasks, inter-processor and inter-process communication. This should be minimized so that a load balancing technique can work efficiently.

Fault Tolerance is the ability of an algorithm to perform uniform load balancing in spite of arbitrary node or link failure. The load balancing should be a good fault-tolerant technique.

Migration time is the time to migrate the jobs or resources from one node to other. It should be minimized in order to enhance the performance of the system.

Response Time is the amount of time taken to respond by a particular load balancing algorithm in a distributed system. This parameter should be minimized.

Scalability is the ability of an algorithm to perform load balancing for a system with any finite number of nodes. This metric should be improved. Performance is used to check the efficiency of the system.

V. PROPOSED SYSTEM

The proposed system mainly focuses on how to improve the energy efficiency of servers through appropriate scheduling strategies. Here a new energy-efficient task scheduling model is proposed by considering the relationship between the performance and energy consumption. To achieve better performance the optimal performance and energy point is required for each server in the cloud environment. An efficient resource management approach is suggested to avoid overload in the system while reducing energy consumption in large scale cloud environments. In this approach, the Dynamic Voltage and Frequency Scaling technique is used to adjust the work frequencies of servers based on the current workload. By optimizing this measure, the utilization of servers can be improved in the face of multidimensional resource constraints. The resource utilization and energy consumption can be minimized in the Cloud environment as a variant of Multidimensional Fractional Knapsack problem which is solved with dynamic programming algorithms.

System Description :

Assuming that there are M servers in a data center, denoted by $A = \{a_1, a_2, \dots, a_M\}$. The current CPU utilization of server a_k is BX_k and its optimal point is BO_k where $k = 1, 2, \dots, M$. The input data D will be divided into y splits with each size of 64 MB, denoted by $D = \{d_1, d_2, \dots, d_m\}$. We randomly store the y splits on M servers. To ensure the reliability each split of data block is stored in three different servers. The element $E_{ij} \in \{1, 2, \dots, N\}$, in which $i = 1, 2, \dots, a$ and $j = 1, 2, 3$.

Optimization model:

The proposed solution considers the resource infrastructure as a pool of elastic resources using any technology for resource virtualization. Therefore the resources are divided in small chunks and allocated for incoming jobs. The amount of given resources can vary based on the workload generated by the job requirements. The problem of maximizing the resource utilization and the profit while minimizing the energy consumption is handled with optimization problem.

In order to conclude to the most well-fitted solution, we also model the profit from job execution. Profit in



the examined case, is a multi-parametric function, depended on the probability of an SLA violation given the amount of committed resources, the compensation cost and the infrastructure operating cost yielding from a job execution. The implementation approach takes into consideration the cost of a job with specific requirements, the cost of an SLA violation and the probability of an SLA violation, given the job's workload in order to calculate the profit at any time. Here the objective of this minimum knapsack is to find the least value such that the total weight of the selected items is at least the capacity. The proposed system also predicts the future workload for managing and scheduling the power demand of large-scale cloud environments. Given a set of jobs to be submitted, each with a specific workload and a desired outcome. Proposed system determines the number of jobs to be executed and the resources to be committed so that the total workload is less than or equal to a maximum limit and the total value is as large as possible.

The Fractional Knapsack problem is a variant of the classic knapsack problem. Given objects of different values per unit volume and maximum amounts and being able to take pieces of objects and find the most valuable mix of objects which fit in a knapsack of fixed capacity. However, resource allocation for the incoming jobs is dynamic and can be fractional. For example, assume that jobs j_1 to j_{i-1} are already running and an incoming job arrives requesting resources. If the new job requires memory and the available memory is less than the amount required. Then only a fraction of the desired amount can be allocated for it, by taking the QoS into considerations.

In a multi-dimensional knapsack, the constraint,

$$\sum_{j=1}^n w_j x_j \leq C_i \quad \forall i \in \{1, \dots, d\}$$

In Cloud computing, a data center can run on several heterogeneous servers. Each server is equipped with different hardware such as the processor, RAM, and the hard disk. For reducing energy consumption, each server can adjust its work frequency according to workloads. Our aim is to minimize the energy consumption and maximize the profit. The task i is assigned to the speed j then the decision variable $x_{ij}=1$ otherwise it takes the value 0.

$$\sum_{i=1}^n \sum_{j \in \Omega_i} P_{ij} X_{ij} \leq C_i$$

subject to

$$\sum_{i=1}^n \sum_{j \in \Omega_i} C_{ij} X_{ij} \leq C$$

$$i=1 \dots n$$

$$\sum_{j \in \Omega_i} X_{ij} = 1 \quad 1 \leq i \leq n$$

There are two processes in our work. First is to provide the feasible combination or scheduling for a job. Second is to provide the appropriate voltage and frequency supply for the servers via the DVFS technique. The requirements of a job are given in terms of maximum and minimum frequencies. The minimum frequency is to ensure the lowest limit of the executing performance for a job. A list of available servers is given by the VM manager. The VM manager has to list the server statuses including used resources and remainder capacities.

The VMs created on a server also have limited capacity according to the resources provided by the server. While allocating a job to proper VMs, we also have to know the resources of the related VMs. We can derive the energy consumption of VMs from the resources used by VMs. In our work, we want to lower the working frequency to reduce the energy consumption of executing jobs. The optimization algorithm for the objective function works as follows.

- Step 1 : Receives new job and the constraints from the user.
- Step 2: Select the appropriate physical machine and Create the required VM based on the requirements of the newly arriving job
- Step 3: Calculate the workload for each VM from the Corresponding servers.
- Step 4: Assign VM to the job based on the availability of the servers in the environment
- Step 5: If the request is not satisfied with the Available VM then the scheduler selects the new server from the availability list.
- Step 6: If there is no server available, the assignment Failed, and goto step 7.
- Step 7: Add the new server into the availability list and goto step 4.
- Step 8: Finalize the process of scheduling by allocating the appropriate assignment of VM to the physical server.
- Step 9: Stop the process.

The concept of skewness can be used to quantify the unevenness in the utilizations of multiple resources on the server. A load prediction mechanism can be employed to capture the future usage accurately. Prediction of workload can be done based on the past external behaviors of virtual machines without migrating the existing virtual machines. This algorithm



can be executed periodically to evaluate the resource allocation status based on the predicted future resource demands of VMs. To estimate the resource requirements of the application to remind the system to create or collect the VMs in accordance with the real Internet situation.

Future Prediction

The prediction technique makes use of existing behavior with respect to the hardware based counter. Since the prediction algorithms plays a major role in the performance of server by improving the stability and the resource allocation decisions.

A server as a hot spot can be defined if the utilization of any of its resources is above a hot threshold. This indicates that the server is overloaded and hence some VMs running on it should be migrated away. The temperature of a hot spot is also defined as the square sum of its resource utilization beyond the hot threshold. It reflects its degree of overload. If a server is not a hot spot, its temperature is zero. A server may be turned into cold spot if the utilizations of all its resources are below a cold threshold. This indicates that the server is mostly idle and a potential candidate to turn off to save energy. However, this can be done only when the average resource utilization of all actively used servers in the system is below a green computing threshold. A server is actively used if it has at least one VM running. Otherwise, it is inactive. Finally, a warm threshold is defined to be a level of resource utilization that is sufficiently high to justify having the server running but not so high as to risk becoming a hot spot in the face of temporary fluctuation of application resource demands. Also, different types of resources can have different thresholds. For example, the hot thresholds for CPU and memory resources can be fixed as 90 and 80 percent, respectively. Thus a server is a hot spot if either its CPU usage is above 90 percent or its memory usage is above 80 percent. The servers are listed in the decreasing order of hot spot temperature. For each server, the decision should be made on which of its VMs should be migrated away. The objective is to migrate away the VM that can reduce the server's temperature the most. In case of ties, the VM whose removal can reduce the skewness of the server the most can be selected. The proposed system aims to eliminate all hot spots if possible. Otherwise, keep their temperature as low as possible.

When the resource utilization of active servers is too low, some of them can be turned off to save energy. The challenge here is to reduce the number of active servers during low load without sacrificing performance either now or in the future so that

oscillation in the system may be avoided. When a new request for a virtual machine arrives, an effective optimization based resource dispatch technique may be adopted to determine the reallocation of resources. This techniques would help to achieve an effective configuration via reaching an agreement between the utilization of resources within physical machines monitored by physical machine monitor and service level agreements (SLA) between virtual machine operator and a cloud service provider.

VI. RESULTS AND DISCUSSIONS

In a simulated environment the results illustrates the differences among different algorithms. The DVFS based algorithm has the higher energy consumption under the light workload because it performs local energy saving by adjusting clock frequency in each computing node without migrating VMs among nodes.

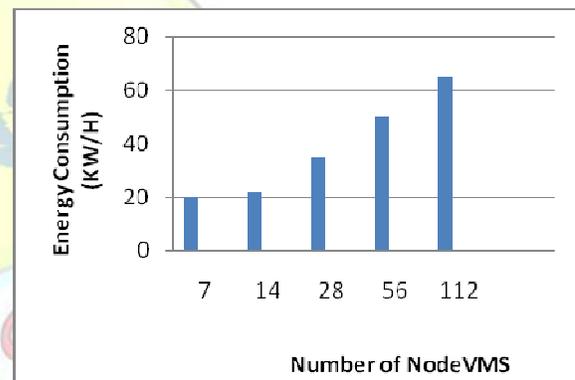


Fig :1 Energy Consumption of DVFS based scheduling

Fig :2 shows the performance of First Fit, Best Fit and DVFS based algorithms. In terms of energy saving the level of energy consumption is high for the DVFS without violating the SLA. Fig :3 shows the scalability of DVFS based algorithm under light and heavy loads. It provides nearly the same linear growth of energy saving capability.

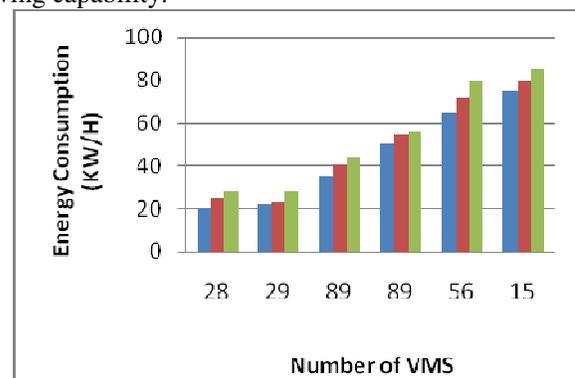


Fig :2 Energy Consumption of different scheduling

Algorithms with varying load

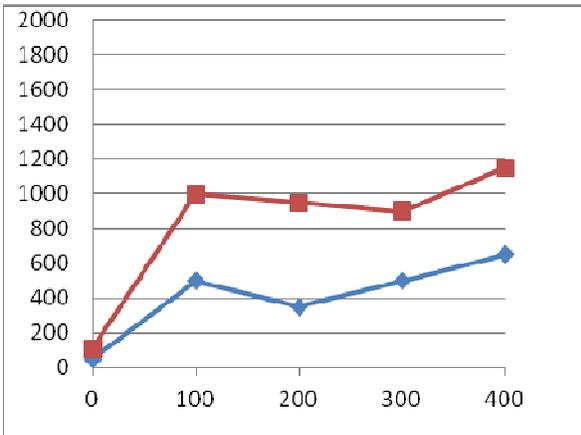


Fig :3 Energy Consumption of DVFS based Scheduling with varying load

VII. CONCLUSION

In the present scenario, there is a need to develop an energy-efficient resource management technique that can improve the performance of cloud computing by balancing the workload across all the nodes in the cloud along with maximum resource utilization, in turn reducing energy consumption and carbon emission to an extent which will help to achieve Green computing. A prediction mechanism is designed to estimate the number of resource utilization according to the SLA of each process and the resources are redistributed based on the current status of all virtual machines installed in physical machines. Based on the new scheme, Our proposed system can fully utilize hardware resources and maintain desirable performance in the cloud environment.

REFERENCES

[1] Xiaoli Wang ,Yuping Wang, Hai Zhu, Energy-efficient task scheduling model based on MapReduce for cloud computing using genetic algorithm: Journal of Computers, Vol.7, NO.12, December 2012

[2] Nidhi Jain Kansal, Inderveer Chana, Existing Load Balancing Techniques in Cloud Computing: A Systematic Review :Journal of Information Systems and Communication Vol 3,Issue 1,2012,pp-87-91.

[3] Zhen Xiao,Weijia Song, and Qi Chen, Dynamic Resource Allocation Using Virtual Machines for Cloud Computing Environment : IEEE Transactions on Parallel and Distributed Systems, Vol 24, NO 6, June 2013.

[4] L.Ceuppens, A. Sardella, D. Kharitonov, Power saving strategies and technologies in network equipment opportunities and challenges, risk and

rewards, in: Proc. of Applications and the Internet 2008, 2008, pp. 381–384.

[5] M. Etoh, T. Ohya, Y. Nakayama, Energy consumption issues on mobile network

systems, in: Proc. of Applications and the Internet 2008, pp. 365–368.

[6] Gartner, Gartner estimates ICT industry accounts for 2 percent of global CO2 emissions, <http://www.gartner.com/it/page.jsp?id=503867>.

[7] Amazon Elastic Compute Cloud (Amazon EC2), <http://aws.amazon.com/ec2,2008>.

[8] M.A. Vouk, Cloud computing-issues, research and implementation, in: Proc.of the ITI 2008 30th Int. Conf. on Information Technology Interfaces, Cavtat, Croatia, June 2008, pp. 31–40.

[9] Anton Beloglazov, Jemal Abawajy, Rajkumar Buyya, Energy-aware resource allocation heuristics for efficient management of data centers for Cloud computing, Future Generation Computer Systems 28 (5) (2012) 755–768.

[10] Nikzad Babaii Rizvandi, Javid Taheri, Albert Y. Zomaya, Young Choon Lee,Linear combinations of DVFS-enabled processor frequencies to modify the energy-aware scheduling algorithms, in: Proc. of the 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, 2010,pp. 388–397.

[11] Christo Ananth, Kavya.S., Karthika.K., Lakshmi Priya.G., Mary Varsha Peter, Priya.M., “CGT Method of Message forwarding”, International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE), Volume 1, Issue 1, August 2015,pp:10-15

[12] Vijindra, Sudhir Senai, Survey on Scheduling Issues in Cloud Computing : International Conference on Modeling Optimization and Computing-Procedia Engineering 38,2012, pp-2881-2888.

[13] A. Lewis, S. Ghosh, N.F. zeng, Run-time energy consumption estimation based on workload in server systems, in: HotPower’08 Proceedings of the 2008 Conference on Power Aware Computing and Systems, 2008, pp. 4–4.

[14] S. Rivoire, P. Ranganathan, C. Kozyrakis, A comparison of high-level full-system power models, in: Proceedings of the 2008 Conference on POWER AWARE COMPUTING and Systems, 2008.

[15] D. Kusic et al., Power and performance management of virtualized computing environments via look ahead control, in: Proc. of 5th IEEE Intl. Conf. on Autonomic Computing (ICAC 2008), Chicago, USA, June 2008, pp. 3–12.