A Study on Virtualization and Task Efficient Scheduling Algorithms in Cloud Environment

Gagandeep Kaur[1], Dr. Gagandeep Jagdev[2]

[1]*Research Scholar (M.Tech.), Yadavindra College of Engineering, Talwandi Sabo (PB)*

[2]*Dept. of Comp. Science, Punjabi University Guru Kashi College, Damdama Sahib (PB)*

*Corresponding Author: Dr. Gagandeep Jagdev, *Dept. of Comp. Science, Punjabi University Guru Kashi College, Damdama Sahib (PB)*

Abstract: *Energy efficiency has been the most vital for future information and communication technologies (ICT). The call for energy-efficient technologies needs to make louder in order to reduce greenhouse gas emissions. Cloud computing has gained considerable attention as a promising approach for delivering ICT services by improving on utilization and consumption of data center resources. Cloud computing focus on hardware aspects needs to be extended to system operation and networking aspects. Today in the field of ICT, businesses, and companies care much about achieving a sustainable strategy for their operations. Cloud computing enables accessing computing resources such as applications, storage, movies, video games, storage and music on demand in such a manner that it is totally abstract to the clients and they have no idea from where they are receiving these contents. Energy efficiency refers to the practice of performing the same task as before but by consuming less energy. The environment is suffering from increasing hazardous energy emissions by large industries.IT sector has a deep impact on our environment along with consuming a notable amount of energy.The central theme of the research paper is to examine the pros and cons of cloud computing from energy efficiency aspect along with analyzing energy efficiency in case of Virtual Machine (VM) scheduling. The paper also discusses the working of three prominent scheduling algorithms, FCFS (First-Come-First-Serve), SJF (Shortest Job First), and RR (Round Robin).*

Keywords: *Cloud Service Provider (CSP), Energy, Power, Virtualization, Virtual Machines (VMs)*

Abbreviations: *CSP, FCFS, IaaS, PaaS, SaaS, SJF, RR, VM*

1. INTRODUCTION

Today it has been made possible that people are consulting their email online through webmail clients, working on collaborative documents using web browsers and creating virtual albums. People are running applications on servers and even storing their crucial data on servers rather than on their systems. All this has been made possible because of cloud computing. The cloud environment is shown in Fig. 1.

Fig1. *The figure depicts the Cloud Computing Environment*

The name cloud computing has been given because of a metaphor about the internet as the user can see the entire system as a cloud but can view what's inside it.

➢ The IT encyclopedia whatis.com explains cloud computing as a general term for anything that involves delivering hosted services over the internet. The three main features that distinguish cloud from traditional hosting are

• It is available to be sold on demand as required, by a minute or by an hour.

• It is elastic as one can use as much service as they desire at any time.

• All this provided service is managed by the provider.

➢ IEEE defines cloud computing as a model in which information is stored permanently on servers on the internet and temporarily cached on clients.

➢ Wikipedia illustrates cloud computing as computation, data access, software and storage services which are not confined to end-user knowledge of the physical location and configuration of the system that delivers services.

So, it can be concluded that cloud computing is a computational model based on consumption of resources, applications, and hardware offered on the internet and used on demand.

1.1. Cloud Computing Actors

Cloud environment comprises three main actors performing different roles and interactions. These actors are mentioned as under.

➢ **Cloud Service Provider (CSP)**

CSP possesses the cloud infrastructure on which services related to the cloud are to be deployed. CSP is responsible for management and control of cloud resources for handling user's requests.

➢ **Cloud User**

An enterprise/company/organization or an individual who consumes the services provided by CSP are referred as cloud user.

➢ **Cloud Broker**

The intermediary between above two mentioned actors is referred as a cloud broker. Cloud Broker is responsible for distribution of incoming requests between different providers based on users' needs and requirements.

1.2. Classic Cloud Service Models

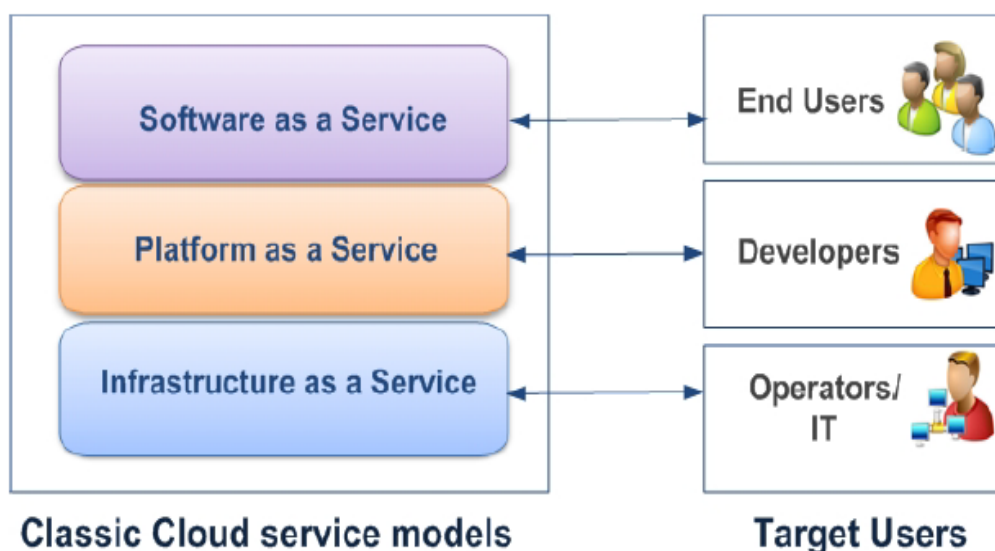Classic cloud service models are categorized as under (Fig. 2).

Fig2. *Service models in Cloud Computing*

➤ **Infrastructure-as-a-Service (IaaS)**

IaaS is a most simple model for delivering services related to cloud. It comprises provisioning and delivering of basic resources like physical servers, virtual machines, network, and storage. It enables companies to rent resources and use them on demand instead of possessing their own resources locally. It enables users to have direct access to the lowest level in the stack and begin building their application environments from scratch. For example, Amazon EC2.

➤ **Platform as a Service (PaaS)**

Compared to Iaas, Paas is more complicated and higher level service. It provides environment for application development and software platforms to develop, deploy and manage cloud applications while not worrying about the kind of technology being used and also hides low-level details from users. For example, Google App Engine and Microsoft Azure Services.

➤ **Software as a Service (SaaS)**

SaaS is the highest level of cloud service model. SaaS provides complete applications to users through the internet. SaaS providers have complete control on application software's and manage infrastructure. Users access the desired applications as if they were placed locally and don't need to have any idea about working on cloud and technologies involved. For example, Google Documents, Google Apps, social media platforms, emails, and project management.

As the cloud computing is enhancing itself in every field and sector, the data centers are becoming large in scale and consume more energy. The CO_2 emissions made by these data centers is further increasing the level of global warming and also the power consumption made by these data centers. According to Gartner Report, average power consumption by a Data Center is so high that it can be used to provide power to 25000 homes. Today every enterprise is shifting itself from traditional to online and further enhancing the demand for servers. To name few of the major Cloud Service Providers (CSP) are Amazon Web Services, Google App Engine, and Microsoft Azure. The new data centers require more energy consumption each year. It is mentionable here that an idle server consumes approximately 70% of its peak power. It means that even if server is not in use, it is consuming 70% of its peak power. In this regard, Intel's focused on the need for dynamic resource scheduling which emphasizes on switching off the idle servers. To minimize the power consumption, the concept of virtualization is used in the cloud computing. The Virtual Machines (VMs) used resembles the physical environment and reduces the need for hardware [6, 7]. The concept is depicted in Fig. 3 where each physical server comprises n number of VMs and Data Centers comprises of n number of physical servers. By managing these VMs and DCs, we can contribute largely in achieving Green Cloud Computing. The purpose of energy efficient algorithms is to shift VMs from servers having fewer loads to servers having larger loads. During different loads, these energy savings are significant. Every enterprise does not have its own cloud, these companies outsource the cloud services from other companies [3, 8].
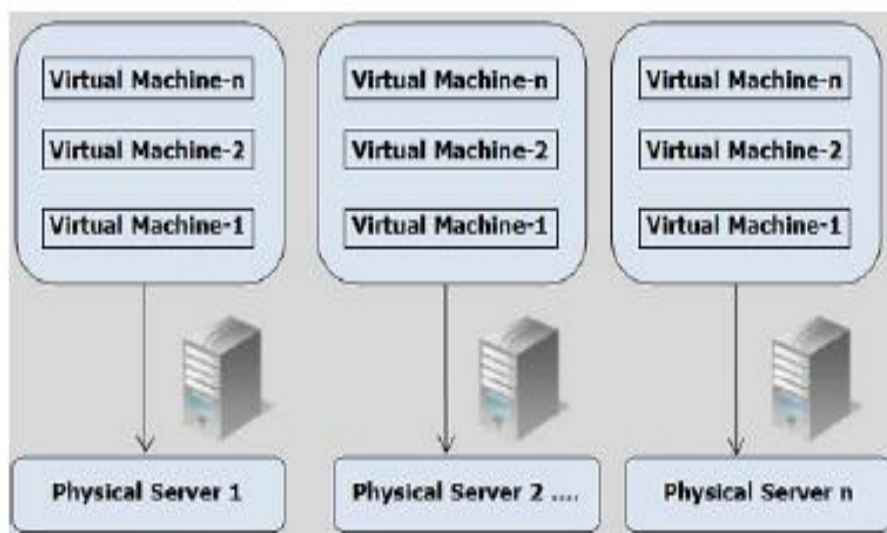

Fig3. *Concept of Virtualization*

It is also necessary that provided solutions in regard with energy efficiency scales in multiple dimensions and CSPs should deal with user requirements which are complex enough. Requested services are more complicated and complete since there is a need for users to deploy their own applications with the topology they choose and having control on both infrastructure and programs. Another important factor is the type of virtualization. CSPs are adopting new container-based virtualization technologies like Docker and LXC that enable the deployment of applications into containers.

Another important dimension to be considered is the resource provisioning plan. CSPs can offer two types of resource provisioning: on-demand and advance or long-term reservation. Advance reservation concept has many advantages for co-allocation of resources. It offers provision for resource planning and reservation in future and offers a promise of providing resources when required. Despite, it advantages, more focus is laid on on-demand plan [9].

2. VIRTUALIZATION AND ITS TYPES IN CLOUD COMPUTING

Virtualization [19, 20] acts as the backbone of cloud computing. It is a kind of abstraction in which several virtual resources are multiplexed on physical one as shown in Fig. 4. The objectives behind utilizing virtualization are to achieve higher resource utilization, easy resource management, isolation, and resource elasticity and to enable heterogeneous services to co-exist on same physical hardware [10].
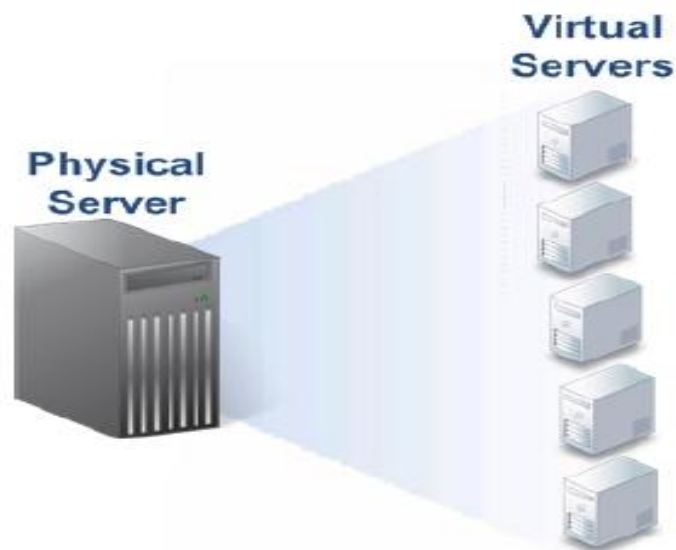

Fig4. *Existence of multiple virtual servers on single physical machine*

2.1. Server Virtualization

Server utilization is the most common resource abstraction technique in Cloud Computing. This virtualization allows multiple isolated virtual servers to execute on a single one and be implemented in multiple ways. Implementation includes full virtualization, para-virtualization, and OS-level virtualization. Full virtualization and para-virtualization makes use of hardware but are different on how the host and guest operating systems are modified to support virtualization and also on how they interact with each other. On the other hand, operating system level virtualization does not use hypervisor at all. In this virtualization, all the virtual servers run same host OS that performs all functions of fully virtualized hypervisor [2, 4, 11].

Server utilization can be classified into two main categories.

➢ **Hypervisor-Based Virtualization**

It is a traditional method of performing virtualization in cloud computing. This technology works on hypervisor which is a layer of software responsible for managing physical server resources. For example, VMWare, Microsoft Hyper-V, Xen and Virtual Box. Guests are referred as Virtual Machines (VMs) and run different operating systems like Linux and Window on top of the same physical host.

OS or Container-based virtualization. If additional software layer is introduced by this kind of virtualization, it enables resource consolidation into virtualized servers and offers live migration feature to move VMs to other servers without any need of shutting them down [12].

➢ Container Based Virtualization

It is a lightweight alternative to hypervisors. It involves technology similar to operating system level which provides provision of running multiple isolated environments on the same host. Containers are based on shared operating systems and unlike conventional VMs, they do not execute on multiple operating systems and makes use of the single operating system. For example, Docker, Linux containers, Solaris Containers, OpenVZ and Virtuozzo Containers. When performance matters, container-based servers are regarded best. It provides most efficient resource usage as it supports a large number of instances on the single host. It also provides portability, transport and process-level isolation across hosts [13, 14].

Although being different, both hypervisor and container-based virtualization are not exclusive but complementary and increasingly used together. Container-based virtualization is building lightweight PaaS environments and hypervisors are used for IaaS cloud services. Complex services can deploy both the applications over hybrid IaaS/PaaS providers. For example, Proxmox has both technologies on same physical layer [18].

3. ENERGY EFFICIENCY IN CLOUD DATA CENTERS

To improve energy efficiency in the cloud environment, it is important to have a deep insight of power distribution and consumption in data centers. It is estimated that server consumes 80% of total IT load and 40% of total data center power consumption. Remaining power is consumed by other devices like distribution wiring, air conditioners, transformers, pumps, and lighting. The big giants like Google, Facebook, eBay has taken steps to reduce power consumption utilized on cooling data centers. In this approach, the air temperature is lowered by using natural cool air or water instead of mechanical refrigeration. In some climates, these savings can reach 100% as zero refrigeration is required [1, 15].

Major causes of energy waste are mentioned as under.

➢ Low Server Utilization

With the increase in a number of data centers, the number of servers is also multiplying. It has been observed that most data center servers are underused. Servers draw between 60 to 90% of peak power while average server utilization is much low ranging near 20%. Consolidating virtual servers on a lesser number of hosts allows running the same applications with much low power consumption.

➢ Idle Power Waste

About 85 to 95% of the time, data servers sit idle and are not processing any useful work. An idle server consumes 70% of its peak power when not in use. So, it is needed to turn off the idle servers when they are not in need.

➢ Lack of Standardized Metric

To optimize energy efficiency, it is important to use energy efficiency metrics for servers to sort them according to energy efficiency and enable scheduling algorithms to make decisions and select the best resource available.

➢ Energy Efficient Solutions are Not Adopted

The big cloud giants do a great job in saving energy and represent less than 5% of the total energy usage. The other 95% are much less energy efficient and involves small, medium, corporate and multi-tenant operations. So, it is needed that small data centers consider this matter seriously and adopt appropriate steps in this regard.

4. POWER MEASUREMENT AND MODELING IN CLOUD COMPUTING

Power and energy have a deep relationship. Power consumption refers to the rate at which machine can perform its work and can be found by multiplying voltage [1, 2]. Power is measured in Watt (W). Energy is referred as an amount of power used over a period of time. Energy is measured in Watt-

hour (Wh). The relation between energy and power is mentioned below where P is power consumption, I is current, V is voltage, T is time interval and E is energy.

$P = I * V$

$E = P * T$

Power metering models estimate the power consumption of servers and VMs using OS-provided metrics or hardware-provided metrics.

➢ **Power Measurement for Servers**

The best method to get accurate information about the energy consumed by servers is to directly measure it. But this needs extra hardware to be installed, add intelligent monitoring capabilities and deal with enormous amount of data. For example, Green Open Cloud (GOC) is one such example of energy monitoring and measurement framework that depends on energy sensors to measure the electricity consumed by Cloud resources.

➢ **Power Measurement for VMS**

Unlike power measurement of servers, power measurement of VMs is not possible and cannot be done by any sensor and neither be connected to any hardware measurement device. Although efforts have been made in this regard by retrieving the idle power from the power consumption of server when it hosts VM, which is not very precise.

Because it is not possible to measure power consumed by VM, the models that estimate the energy consumption and VM migration power cost are more attractive for power metering. As the CPU of a server consumes the most important amount of power and also the relationship between power and CPU utilization is linear, CPU based linear models represent a lightweight and simple way to estimate server's power usage. Many simple utilization power models for servers have been proposed. All these assume that CPU is the only factor in their power models and present an approximation for total power against CPU utilization (Fig. 5).

$P = P_{idle} + U * (P_{peak} - P_{idle})$

P is total power consumption, P_{Peak} is peak power consumption, P_{idle} is idle powerconsumption, and U is CPU utilization (a fraction between 0 and 1).
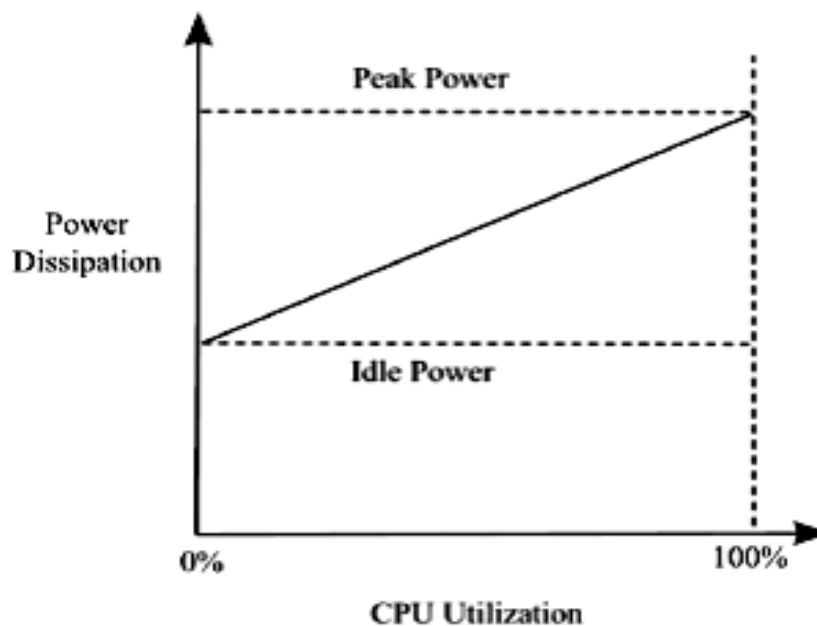

Fig5. *Figure depicts relationship between CPU utilization and Power dissipation*

5. RESOURCE ALLOCATION IN CLOUD COMPUTING

Resource scheduling [16, 17] is the most important tasks in cloud computing. It involves identifying and assigning resources to every user request in such a manner that user requirements are met and

goals of CSP are satisfied. These goals are primarily related to energy consumption or cost optimizing. The wider adoption of cloud computing and virtualization technologies has led to tremendous increase in cluster sizes ranging to hundreds in case of mini data centers and to thousands in case of large data centers. This has resulted in high demand for electricity consumption, enhancing data center ownership and maintenance cost further. It also impacted the environment by increasing carbon footprints. It is because of all this; energy efficiency is turned to be a matter of concern in cloud computing. Fig. 6 depicts the resource allocation concept involved in cloud computing.
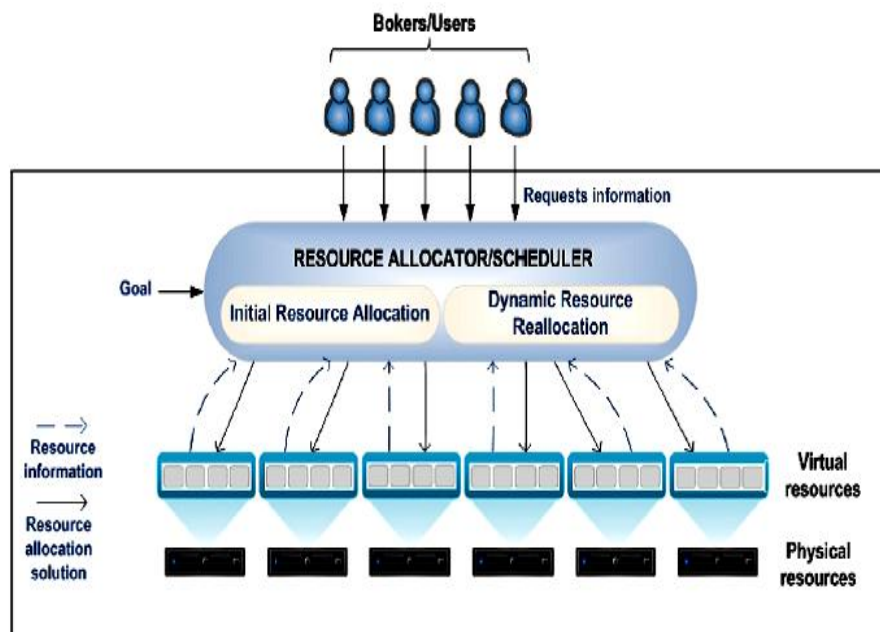

Fig6. *Resource Allocation in Cloud Computing*

CSP can offer different types of provisioning plans. Two most commonly known are on-demand and advanced reservation plans[5].

➢ **On-Demand Resource Allocation**

Majority of CSPs depend on on-demand policy to allocate resources. In this kind of policy, the request is accepted if the resource is available or else it is simply rejected. The stress is laid on proposing energy aware heuristic algorithms and policies in order to save energy by reducing the number of running servers. The central idea is to consolidate tasks and applications on a minimum number of servers and switch off surplus machines. In this regard, a nature-inspired VM consolidation algorithm resembling Ant Colony optimization has also been proposed whose aim is to reduce the number of physical machines and save energy.

➢ **Advanced Resource Reservation**

This policy provides a feature for resource planning in future and offers increased expectation that resources can be provided when demanded. Besides this provision, CSP still prefers simple resource allocation policies like on-demand.

6. SCHEDULING ALGORITHMS

The algorithm describing the working of three most prominent scheduling algorithms are mentioned as under. Below is a detailed explanation of what each code does.

FCFS (First Come First Serve)

1. The first come first scheduling algorithm executes the process that comes first, finishes it and then take on the next process in the queue.

2. The burst times of processes are hard coded in the array 'btime'.

3. Then the wait time of each process can be found by adding the waiting time and burst time of process preceding it.

4. These values are stored in an array 'wtime'. The total waiting time 't1' can be found by adding the waiting times of individual processes.

5. The turnaround time is found using adding the burst time and waiting time of each process and it is stored in the array 'tatime'.

6. 't2' is the total turnaround time found by adding individual turnaround times.

SJF (Shortest Job First)

1. The shortest job first algorithm runs the process whose burst time is lowest.

2. We assumethat all the processes have already arrived.

3. The burst times are hard coded in the array 'ptime'.

4. As all the processes have arrived and burst times are known they are sorted using appropriate loops.

5. Then the wait time of each process can be found by adding the waiting time and burst time of process preceding it.

6. These values are stored in an array 'wtime'.The total waiting time 'total' can be found by adding the waiting times of individual processes.

7. The turnaround time is found using adding the burst time and waiting time of each process and it is stored in the array 'tatime'.

8. 'total2' is the total turnaround time found by adding individual turnaround times.

RR (Round Robin)

1. The round robin algorithm runs a process for a given quantum time. If the process gets completed in the given time it is dequeued else it is put at the end of the queue so that it can run again later.

2. The burst time and quantum time are hard coded in array 'btime' and variable 'q' respectively.

3. Initially the remaining time array 'rtime' is equated to the burst time as no process has run yet.

4. A process which is present in the starting of the array is selected and its remaining time is compared with the quantum time.

5. If remaining time (rtime) of process is greater than the selected process is run for quantum time(q) and this q is subtracted from rtime. For other processes q is added to waiting time which is an array 'wtime'.

6. If quantum time is greater the process is run to completion and rtime is set to 0 for the process selected. For other processes q is added to waiting time.

7. There is also a 'flag' variable which is set if the rtime of one or more processes is greater than zero.

8. So if the flag is set the 4, 5, 6 steps are repeated.

9. The turnaround time is found using adding the burst time and waiting time of each process and it is stored in the array 'tatime'.

10. The variables 'b' and 't' store values of total waiting time and total turnaround time which is found by summing values of all the processes.

7. CONCLUSION

The paper has analyzed and reviewed the potential impact of energy saving strategies for managing an integrated system that comprises a computer system and networks and discussed the three scheduling algorithms (FCFS, SJF, RR). There are few major components of cloud architecture which are responsible for the high amount of power dissipation in the cloud. The paper proposed following concepts in brief.

➢ To identify the main sources of energy consumption and noteworthy trade-offs between performance and energy efficiency.

➢ Surveyed how energy savings can be achieved in large-scale computer services that integrate communication needs.

REFERENCES

[1] Arindam Banerjee et al., "Energy Efficiency Model for Cloud Computing", *IJEIC*, Vol. 4, Issue 6, 2013, pp. 29-42.

[2] Erol Gelenbe et al., "Energy-Efficient Cloud Computing", DOI: 10.1093/comjnl/bxp080, Oxford University Press, August 2009.

[3] Chaima Ghribi et al., "Energy Efficient VM scheduling for Cloud Data Centers: Exact allocation and migration algorithms", *International Symposium on Cluster, Cloud and Grid Computing (CCGrid), 2013 13th IEEE/ACM,* May 2013.

[4] Anuj Prasher et al., "Areview on Energy Efficient Cloud Computing Algorithms", *IJAIEM,* ISSN2319-4847, Vol. 3, Issue 4, April 2014.

[5] Puneet Himthani et al., "Comparative Analysis of VM scheduling Algorithms in CloudEnvironment", *International Journal of Computer Applications*, Volume 120, N0. 6, June 2015.

[6] Ellendula Madhukare et al., "Efficient Scheduling Algorithm for Cloud", Procedia ComputerScience 50 (2015), pp. 353-356.

[7] Tian, Zhao, Xu, Zhung & Sun "A Toolkit for Modeling and Simulation of Real Time Virtual Machine Allocation in a Cloud Data Center", IEEE Transactions on Automated Science & Engineering, 2013.

[8] Kaur & Kinger "A Survey of Resource Scheduling Algorithms in Green Computing", International Journal of Computer Science & Information Technologies, 4886 – 4890, No. 4, Vol. 5, 2014 [0975 - 9646].

[9] Himani & Sidhu "Comparative Analysis of Scheduling Algorithms of Cloudsim in Cloud Computing", International Journal of Computer Applications, No. 16, Vol. 97, July 2014 [0975 - 8887].

[10] Sonam Rathore "Efficient Allocation of Virtual Machine in Cloud Computing Environment", International Journal of Computer Science & Informatics, 59 – 62, Issue – 3, Vol. 2, 2012 [2231 – 5292].

[11] Agarwal & Jain "Efficient Optimal Algorithm for Task Scheduling in Cloud Computing Environment", International Journal of Computer Trends and Technology, 344 – 349, No. 7, Vol. 9, March 2014.

[12] Deore & Patil "Energy Efficient Job Scheduling & Allocation Scheme for Virtual Machines in Private Clouds", International Journal of Applied Information Systems, No. 1, Vol. 5, January 2013 [2249 - 0868].

[13] Khan, Kapgate & Prasad "A Review on Virtual Machine Management Techniques & Scheduling in Cloud Computing", International Journal of Advanced Research in Computer Science & Software Engineering, 838 – 845, Issue 12, Vol. 3, December 2013 [2277 – 128X].

[14] Wu, Liu, Ni, Yuan & Yang "A Market Oriented Hierarchical Scheduling Strategy in Cloud Workflow Systems", Springer, March 2011.

[15] Li, Peng & Zhang "A Scheduling Algorithm for Private Clouds", Journal of Convergence Information Technology, No. 7, Vol. 6, July 2011.

[16] Gu, Hu, Zhao & Sun "A New Resource Scheduling Strategy Based on Genetic Algorithm in Cloud Computing Environment", Journal of Computers, Academy Publishers, 42 – 52, No. 1, Vol. 7, January 2012.

[17] Soni & Pahadiya "An Energy Efficient Cloud Computing using VM Scheduling Algorithm", International Journal of Research in Engineering, Technology & Management, Issue 4, Vol. 2, July 2014 [2347 - 7539].

[18] Mazandarani & Mormeni "QoS Aware Scientific Application Scheduling Algorithm in Cloud Environment", IISTE Journal of Computer Engineering & Intelligent Systems, No. 12, Vol. 4, 2013 [2222 - 1719].

[19] Choudhary & Peddoju "A Dynamic Optimization Algorithm for Task Scheduling in Cloud Environment", International Journal of Engineering Research & Applications, 2564 – 2568, Issue 3, Vol. 2, May – June 2012 [2248 - 9622].

[20] James & Verma "Efficient VM Load Balancing Algorithm for a Cloud Computing Environment", International Journal on Computer Science & Engineering, 1658 – 1663, No. 9, Vol. 4, Spetember 2012 [0975 - 3397].

AUTHOR'S BIOGRAPHY

Dr. Gagandeep Jagdev, is a faculty member in Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB). His total teaching experience is above 10 years and has above 108 international and national publications in reputed journals and conferences to his credit. He is also a member of editorial board of several international peer-reviewed journals and has been active Technical Program Committee member of several international and national conferences conducted by renowned universities and academic institutions. His field of expertise is Big Data, ANN, Biometrics, RFID, Cloud Computing, Cryptography, and VANETS.