

An Efficient Cloud Scheduling Algorithm for the Conservation of Energy through Broadcasting

Kavita Arjun Sultanpure¹, Abhishek Gupta², L. S. S. Reddy³

^{1,3}K L University, India

²Cloud Security Architect, Intel Corporation, California, USA

Article Info

Article history:

Received Nov 29, 2017

Revised Jun 29, 2017

Accepted Jul 13, 2017

Keyword:

Broadcasting
Cache memory
Caching
Cloud computing
Cloud models
Cloudlet
Energy consumption
Mobile users
Optimization
Scheduling
Server load
Workload

ABSTRACT

Method of broadcasting is the well known operation that is used for providing support to different computing protocols in cloud computing. Attaining energy efficiency is one of the prominent challenges, that is quite significant in the scheduling process that is used in cloud computing as, there are fixed limits that have to be met by the system. In this research paper, we are particularly focusing on the cloud server maintenance and scheduling process and to do so, we are using the interactive broadcasting energy efficient computing technique along with the cloud computing server. Additionally, the remote host machines used for cloud services are dissipating more power and with that they are consuming more and more energy. The effect of the power consumption is one of the main factors for determining the cost of the computing resources. With the idea of using the avoidance technology for assigning the data center resources that dynamically depend on the application demands and supports the cloud computing with the optimization of the servers in use.

Copyright © 2018 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Kavita Arjun Sultanpure,
Departement of Computer Science and Engineering,
K L University,
Green Fields, Vaddeswaram, Guntur, Andhra Pradesh 522502, India.
Email: kavita.sultanpure@gmail.com

1. INTRODUCTION

Cloud computing is included in one of the fastest growing technologies used for the purpose of computing these days. Normally when it comes to cloud computing the models that includes are; Software-as-Service, Infrastructure-as Service and Platform-as-Service. The main focus is about allocating the data and computing it over a scalable information center of the network. Cloud Computing is one of the technology that relies on the allocation of the computing resources rather than relying on local servers and applications. The main purpose behind the designing of the computing technique is to perform efficient cost estimation, application hosting and substance storage. The main properties of the cloud computing technology include:

- Supplying the services to end users quickly using interface on Internet.
- Assigning and organizing the physical resources effectively as well as rapidly.
- Confirming the role of on demand distribution.
- Dealing with a huge quantity of requests and examining needs accordingly.

Cloud computing has been attracting a lot of researchers since last couple of years and is still a great area of research for the keen researchers. Along with the increase in demand and enhancement in the

technology, the complexity of it is widely increasing. There is always one centralized server within a specific area that picks up the services from the sub-servers, and further provides the services to the users. As the cloud network is paid platform, the users expect to get flawless services.

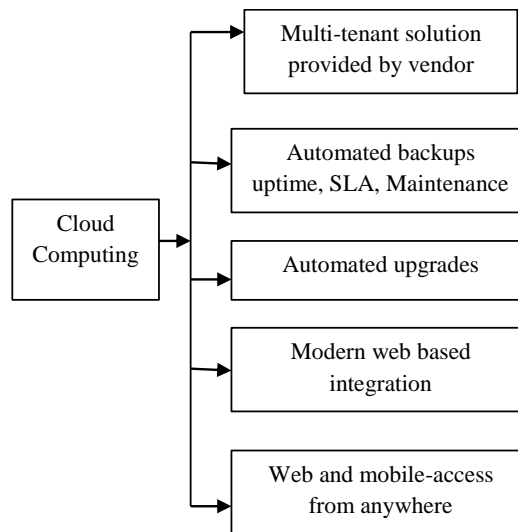


Figure 1. Cloud Computing

Cloud computing is the innovation that utilizes the web as well as the focal remote servers for keeping up the information and applications as shown in Figure 1. With cloud computing the buyers and the organization can use the applications without any establishment and can even access their documents at any system having web access. This innovation takes into consideration considerably more effective registering by bringing together information handling and data transfer capacity.

A straightforward case of cloud computing can be seen in Yahoo e-mail, Gmail, or Hotmail and so on. All you need is only a web association and you can begin sending messages. The server and e-mail administration software are all on the cloud (web) and is completely overseen by the cloud administration supplier including Yahoo, Google and so forth. The customer gets the chance to utilize the product alone and appreciate their advantages.

In order to follow the procedure, the central servers always have to broadcast the requirement which encompasses a huge amount of energy, each and every time.

The paper is divided into five sections. Starting off with the Introduction of the research, described above and following the introduction, we have discussed related work with problem formulation and contribution. Third section includes description of methodology along with the architecture of proposed solution wherein in fifth section of the paper there is a depiction of simulation results.

As stated by NIST, the cloud framework is made up of five important features including: On-Demand Self Service, Broad Network Access, Resource Pooling, Rapid Elasticity and Measured Services.

- a. On-demand self-service: The users are provided by the one-sidedly computing abilities like, server time interval and network storage that is required.
- b. A Broad network access: Competencies are obtainable through the network as well as handled by using customary mechanisms which also stimulates by means of heterogeneous thin or thick user stages (for instances, mobiles, laptops, tablets, as well as work places).
- c. Resource pooling: The supplier's computing assets are assembled to assist numerous users utilizing a multi-tenant framework through dissimilar physical as well as simulated resources, which are vigorously allocated as well as reallocated as per the user's request. There is an intellect of position and individuality in which, the users general has no control or familiarity over the particular position of the offered assets, nevertheless this might be competent to specify position with the side of a greater level of abstraction (for examples, state, country, or information center). Instances of the assets comprise of memory, storage handling, as well as network bandwidth.
- d. Rapid elasticity: Although the competencies are elastically delivered and are unconfined but, still, there are some cases in which the measurement is done with outward as well as inwards proportionate via

request. To the user, the proficiencies obtainable for provisioning frequently, seems to be unrestricted as well as could be taken in any kind of magnitude at any time.

- e. Measured service: Cloud frameworks mechanically control and enhance resource utilization through leveraging a metering proficiency on round about some particular levels of abstraction, which is suitable to the category of services (for example, handling, storage, bandwidth, as well as vigorous client accounts). Asset usage could probably be examined, handled, as well as testified, provided that placidity meant for both of the suppliers and customers of the utilized services.

2. RELATED WORK

A load of research has been done in the field of cloud computing for providing the best computing resources as services and still is providing a great scope for research for the keen researchers. This paper focuses on the broadcasting issue related in the cloud computing and along with that also focuses on the cache concept for the purpose of reducing the energy consumption. The main areas that focus in this work are, memory management, complexities as well as all the basic services. Following is the description of the works that are already done under this topic by different researchers.

Wood et.al [1], presented the issue of security in cloud computing. In this paper, security model for SaaS service has been analyzed by the author. Wang et.al [2], proposed the cost model in cloud computing based on the complexity-aware dynamic task optimal scheduling algorithm wherein Sandeep Kaur [3] proposed the technique based on virtualization for solving the problem of starvation during VM migration. To reduce the memory size, virtualization is a one of the good concepts. From result simulations presented in this paper by author, proposed scheme comes out to be better than previous methods. Qingling et.al [4], analyzed three different types of workloads based on virtualization. The main purpose of the author is to prevent the network from poor performance, associating less memory allocation and maintaining reasonable performance. Large number of VM (virtual machine) will require large memory and hence, low performance is considered while less number of VM gets to deliver high performance. Dimpi et.al [5], presented the survey on various services provided in cloud computing. This paper is just the startup for selection of various services. Cloud computing is based on the virtualized IT resources where different users can share the resources. It has been shown that, when user shares their resources for different services then monthly fees gets availed for this service. Yizeng Chen et.al [6], [7], proposed the comparative analysis of traditional methods with new designed cloud computing based methods for software development. It has been seen that cloud computing is providing good applications for IT industry in terms of SaaS, IaaS and PaaS services [8-20].

Weqing et.al [21], presented various platforms of the cloud computing and also introduced the google technology in cloud computing. It introduced various services like mapreduce, bigtable etc. These services are studied over three clouds. It has also been studied that cloud computing is providing challenges to country as well as industries. Sh Hengliang et.al [22], proposed skyline algorithm for complexity reduction. The main aim is to find the most appropriate node machines.

3. PROBLEM FORMULATION AND CONTRIBUTION

As discussed above, the broadcast mechanism consumes a lot of energy which increases the cost on user end. This paper focuses on developing a cache memory in which the types of services along with the service providers are stored so that for similar type of file request, the server ID is prevented from losing a bulk amount of energy. It looks as a simple solution to the problem but it is a difficult task. Assume a central server with approximately twenty thousand users and hundreds of sub servers. Managing a cache with these many user bases is quite difficult. For mapping directly to the sub server, it is also required to manage the load of the every sub server. If a sub server is providing good response to the posted queries, we cannot route every query to the sub server. The future work of the proposed solution may add an efficient load balancing mechanism for the sub servers for energy saving and for removing idealism to the architecture.

Energy savings are mostly motivated by improved efficiency of data centers while using cloud services like e-mail, calendars, and more. The cloud has multiple number of products at a time, thus, it can send the resources amongst a lot of users more capably that too, with less energy consumption.

- a. User won't have to pay for someone (or a team of someone's) to do things such as install and update software, install and manage email servers and/or fine servers, run backups – the beauty of green computing is that all of the business, related to maintaining the service or application is the responsibility of the green vendor, not yours.
- b. Reduction of carbon footprint: carbon ink usage and papers for print can be reduced by using green computing.

- c. You no longer have to buy software. Besides the convenience of not having to buy software programs and installation process on your own servers/computers, using green applications instead has been proved to be cheaper.
- d. One may be able to consolidate separate application needs into one multi-application green computing service. For instance, Google Apps for business includes e-mail, calendar scheduling application, Google Docs for creating documents, presentations and forms and using online file storage and Google Sites for creating websites, all for only \$5/month for each person on your account. Now think, about the price of, let's say, Microsoft Office (including Microsoft Outlook for email) – and note that Office doesn't include a website application. Green computing vendors such as Info street provides a suite of green application including CRM, calendar scheduling, email, conference calling, file sharing and an employee directory for as little as \$10 per person per month.
- e. Cloud computing helps to cut back on system hardware. File storage, data backup and software programs all take up a lot of space on servers/computers. With green computing, you can use someone else's server to store all this data instead hence, freeing up your in-house computer equipment for other purposes or even letting you get rid of some of it.
- f. A green computing application may make integration easier. As many green computing applications includes an application programming interface (API) you may be able to find it as "compatible" application rather than having to pay to have the applications you want to be integrated and customized for you.
- g. Green computing applications are regularly updated, so you don't have to spend time and money doing it – and giving you the advantage of always having access to an application's latest features and functions.
- h. Green computing allows you and your employee's easy access to applications and data from different computers and devices. "As more consumers and businesses adopt tools such as smart phones and tablets, the ability to host data in the green computing application and access it from just about anywhere on the planet is quickly becoming vital", says Omar El Akkad in Outsource IT Headaches to the Green (The Globe and Mail).
- i. Green computing lets you start up or grow your small business quickly. It's a lot easier and faster to sign up for a green computing application than to buy a server, get it up and running and install software on it. And because you don't need to buy hardware and software, your start up or expansion is cheaper, too.

Factors Affecting Green Cloud Computing:

- a. Rapid Growth of Internet
Dependency on the internet leads to the greater growth in the size and the number of centers. Internet usage is increasing by 10 % due to downloading of the videos and songs. In addition, business has also been using the Internet at high pace.
- b. Increasing Power Density
As CPU consumption has been decreased due to the adoption of the new high tech servers and overall energy consumption is increasing day by day due to high power utilization.
- c. Increasing Cooling Requirement
The increase in the server power consumption leads to the need of high cooling requirement. The ratio of cooling power to server power requirements will continue to increase as data center server densities increases.
- d. Increasing Energy Costs
Data center's expenditure for cooling and power can exceed up to \$ 400 for 500 watts.
- e. Low Server Utilization Rates
Data center efficiency is a major problem in terms of energy use. The server utilization rates are average 5-10 per cent for large data centers.

4. WORKING METHODOLOGY

Cloud server scheduling and maintenance is one of the most difficult and complicated tasks in any cloud server. There are several tasks at any cloud server which need frequent attention. The cloud server doesn't specify any caching for the processes as every bit of the memory is paid at the cloud. Let us consider a situation, in which a file is fetched from the inventory would console around 3J of energy and if, kept into a section from where the file can be fetched in no time then, it may console a slight more energy as compared to the inventory retrieval, it is better to keep the file into cache part for quicker access. No user is ready to wait for so long and hence, quick retrieval of the file and process has become a great necessity in this time frame. The proposed algorithm also provides an opportunity to those users who have downloaded files in bulk and even they can act as a sub server.

Let us consider a situation in which there are 50 users in the network and there are 1000 files. If a specific user has more than 500 files with him, then the central server uses the user as sub server to reduce the burden on the central server. The proposed algorithm has been divided into three sections namely the broadcasting, the caching and third is the server creation. Table 1 shows the server specifications.

Table 1. Server Specifications

Cloud Name	XEN cloud
Total RAM	3024 MB
Total width of coverage	1000 meters
Total height of coverage	1000 meters
Central server location	[500, 500]

M=Total Physical Machine in the Network

Let there be, N number of job owners [A job owner is a user who has some task to be completed at server end]. Each user will have some X and Y location. Each physical machine M_i will also possess some X & Y co-ordinates. Here, in the proposed algorithm we are considering the physical machine as the sub server which would execute the job file request from the user.

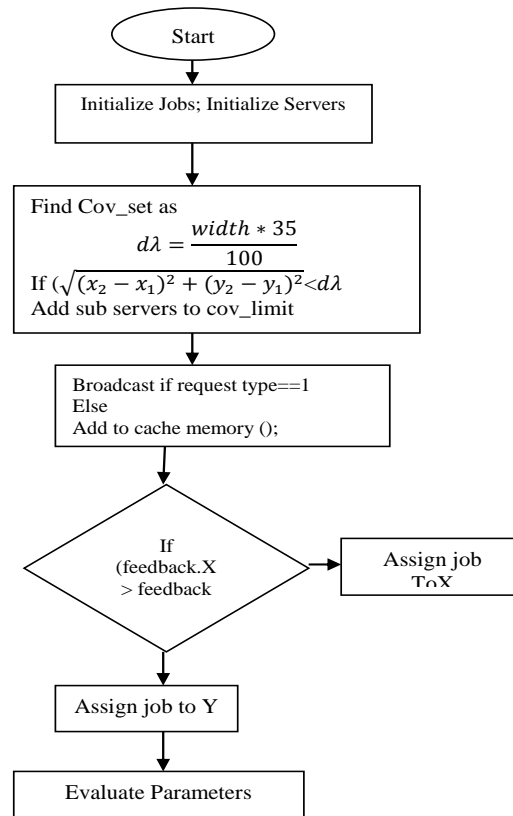


Figure 2. Proposed Methodology

The proposed algorithm shown in Figure 2 first of all will find the allocation probability. The allocation probability determines that which job is going to be executed and through which sub server. For one job, there may be more than one sub server and it is also the possibility that a job gets no physical machine or sub server to get executed. The mathematical representation of the allocation probability is as follows.

Function find allocation prob=(X_a, Y_a, X_{aa}, Y_{aa}) // X_a, Y_a are sub server physical locations and X_{aa}, Y_{aa} are users physical location.

Cov-limit= (width*35)/100

```

for l=1: total physical machine
for j=1: job user count
if ( $\sqrt{((X_{ai}-X_{aaj}))^2 + ((X_{ai}-X_{aaj}))^2} \leq \text{Cov-limit}$ )
Add Yai_user_Id to coverage limit of physical sub server Id based on [Xa, Ya].
end (if)
end (for)
end (for)

```

The mathematical algorithm evaluates that if the distance between the user location (who has a job to be executed) and the physical sub server is less than or equal to 35% of the width of the area, then the probability of the physical sub server to execute the job is high else it is low.

4.1. Broadcasting of the Requirement

The central server will broadcast the requirement of the user if the file has been demanded for the first time and there is no information of the job or file into the cache memory of the network. The central server will keep the responses from the physical machine and will assign the job to the sub server as per their responses and the minimum costing to complete the job. The mathematical expression is as follows.

```

functionBroadcastandAssign = Coverage_Region(Demanding_User, Cov_Set )
foreachsubincov_set(Demanding_User)
Ping_Subserver();
t = Timetowait
ifresponse_time_sub_server <= t
Addsub_serverIdtoPingedbacklist;
end (if)
end (for)
foreacheleinPingedbacklist
cost(ele) = (Hdd_cost*timeofexecution(job)) + ((Ram_cost) *timeofexecution(job))
end (for)
k = find(cost.min)
Assignjobtok
End

```

4.2. Caching

The caching concept makes the execution quite fast and handy. As explained in the broadcast section that the job was assigned to the Kth sub server. Now, this information will be stored into the cache memory and when so ever any other with same kind of file demand or a job with same frequency occurs in the Kth region, then, after the feedback evaluation, the job can be directly assigned to Kth server without wasting the energy of the broadcast.

```

CacheMemory = [0];
Cachecount = 1;
if(RequirementAnalysis == 1)// Requirement of the file has been made for the first time.
req = Requirement;
K = Broadcast.Requirement(req);

```

```

suitable = fingbest(k);
CacheMemory(cachecount,1) = suitable;
CacheMemory(cachecount,2) = Requirement;
Cachecount = 1;
Endif

```

If, the requirement is analyzed as 1 then, such requirement are demanded for the first time then there would be a broadcast. The broadcast will return the response of the sub servers out of which the best suitable server would be selected based on the cost efficiency and the file type and the sub servers details would be stored in the cache memory for further detailing.

4.3. The Feedback Terminology

The Feedback terminology helps the central server to choose the best sub server for any job. After the broadcast phase when the sub server information is stored in the cache, the central server would put a feedback based on the completion time and energy consumed in the completion of the job. If there are more than one sub server for the same job, then based on the feedback it is easy to select which sub server would be best for the job.

$$feedback_{rating} = \sum_{j=1}^2 \left(\sum_{i=1}^n \frac{feedback}{N} + cfeedback \right) / 2$$

$cfeedback = feedback \text{ of the sub servers for the current job}$
 $feedback = All \text{ feedback for } N \text{ jobs}$

5. EXPERIMENTAL RESULTS

This section explains the experimental results by considering the number of parameters that are delay, CPU mis-utilization, energy consumption and delay with respect to the number of users. The comparison of all these parameters is computed with the number of users.

Figure 3 shows comparison for delay between with broadcasting and without broadcasting techniques. The red line shows the delay for the proposed system without broadcasting and the blue line shows the delay for proposed system with broadcasting. Delay in case of with broadcasting is less as compared to the without broadcasting technique. Average delay for with broadcasting is 42ms and average delay for without broadcasting is 36ms.

Figure 4 show the comparison for CPU mis-utilization between with broadcasting and without broadcasting. The red line shows the CPU utilization for proposed system without broadcasting and blue line shows the CPU utilization for proposed system with broadcasting. CPU mis-utilization in case of with broadcasting is less as compare to the without broadcasting technique. Average CPU mis-utilization for with broadcasting is 45% and average delay for without broadcasting is 32%.

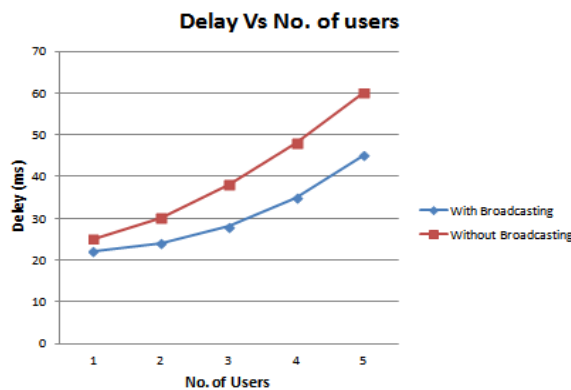


Figure 3. Delay Vs No. of Users with and without broadcasting

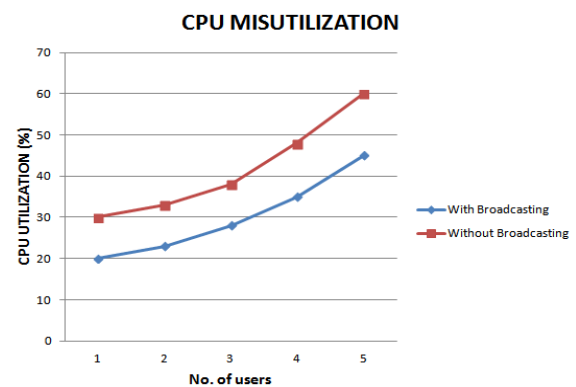


Figure 4. CPU mis-utilization with and without broadcasting

Figure 5 show the comparison for Energy Consumption between with broadcasting and without broadcasting. The red line shows the Energy Consumption for proposed system without broadcasting and blue line shows the Energy Consumption for proposed system with broadcasting. Energy Consumption in case of with broadcasting is less as compare to the without broadcasting technique. We observe that, the average Energy Consumption for with broadcasting is 2.4J and average delay for without broadcasting is 3.2J.

Figure 6 shows the comparison for Overall delay between with broadcasting and without broadcasting by number of users. The red line shows the overall delay for proposed system without broadcasting and blue line shows the overall delay for proposed system with broadcasting. Overall delay in case of with broadcasting is less as compare to the without broadcasting technique. Average overall delay for with broadcasting is 118ms and average delay for without broadcasting is 140ms.

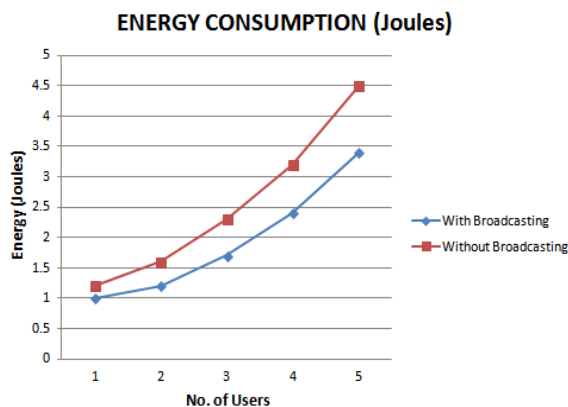


Figure 5. Energy Consumption with and without broadcasting

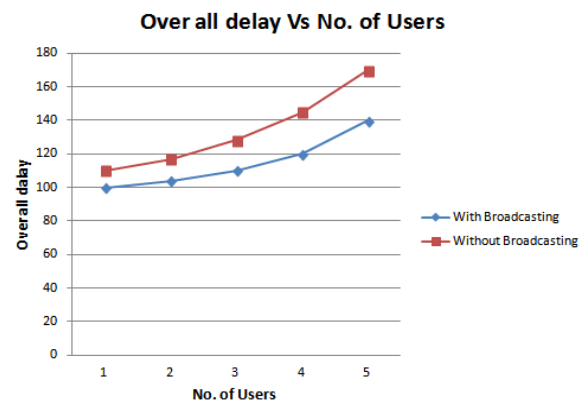


Figure 6. Overall Delay Vs No. of Users with and without broadcasting

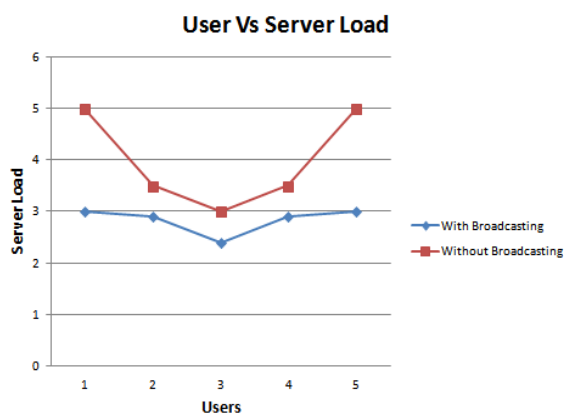


Figure 7. User Vs Server Load with and without broadcasting

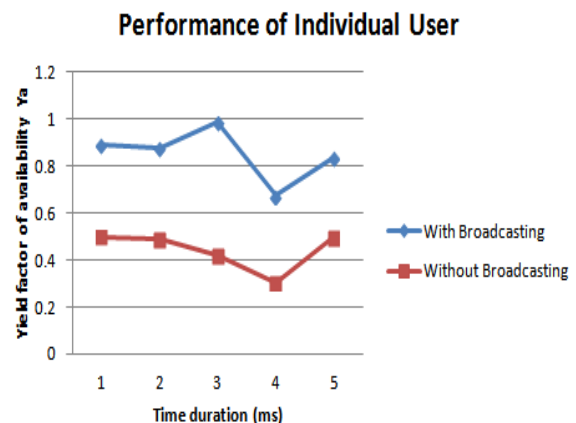


Figure 8. Individual user Performance with and without broadcasting

Figure 7 shows the comparison for Server Load between with broadcasting and without broadcasting with respect to Users. The red line shows the Server Load for proposed system without broadcasting and blue line shows the Server Load for proposed system with broadcasting. Server Load in case of with broadcasting is less as compare to the without broadcasting technique. Average overall delay for with broadcasting is 2.8 and average delay for without broadcasting is 4.

Above Figure 8 shows the comparison for Yield factor between with broadcasting and without broadcasting with respect to Number of cloudlets. The red line shows the Yield factor for proposed system without broadcasting and blue line shows the Yield factor for proposed system with broadcasting. Yield

factor in case of with broadcasting is less as compare to the without broadcasting technique. Average overall delay for with broadcasting is 1.8 and average delay for without broadcasting is 0.4.

6. CONCLUSION

One of the major tasks of the cloud computing is the management of tasks that contribute to the scheduling process. In normal situations only manual decisions were required for processing. But, these methods are not sufficient because of the arrival of the request in the random manner. Hence, it becomes essential to have a scheduling method for managing the tasks. The proposed algorithm is based on the concept of broadcasting with cache concept to get, the request at faster pace. From result simulations, it has been concluded that proposed technique worked well for managing tasks in terms of various parameters like delay = 140 load= 43, energy consumption = 3.5 and CPU mis-utilization = 45 with broadcasting.

REFERENCES

- [1] Wood, Katie, and Mark Anderson, "Understanding the complexity surrounding multitenancy in cloud computing," e-Business Engineering (ICEBE), 2011 IEEE 8th International Conference on. IEEE, 2011.
- [2] Wang ning, "A task scheduling algorithm based on QOS and Complexity-aware optimization in cloud Computing," 2013, vol 5, no 5.
- [3] Kaur, Sandeep, "Memory Management and Reuse Mechanism for Virtual Machine in Cloud Computing to Minimize Energy Consumption: A Review Paper," *Memory*, vol 6, no 6, 2015.
- [4] Wang, Qingling, and Carlos A. Varela, "Impact of cloud computing virtualization strategies on workloads' performance," Utility and Cloud Computing (UCC), 2011 Fourth IEEE International Conference on. IEEE, 2011.
- [5] Rani, Dimpi, and Rajiv Kumar Ranjan, "A comparative study of SaaS, PaaS and IaaS in cloud computing," *International Journal of Advanced Research in Computer Science and Software Engineering*, 4.6 (2014): 458-461.
- [6] Chen, Yizeng, Xingui Li, and Fangning Chen, "Overview and analysis of cloud computing research and application," E-Business and E-Government (ICEE), 2011 International Conference on. IEEE, 2011.
- [7] Liao Li, Zhang Tao, "The Development of Cloud Computing," *Information Technology*, 2010, pp 86-93.
- [8] Chen Kang, Zheng Wei-Min, "Cloud Computing: System Instances and Current Research," *Journal of Software*, vol. 5, 2009, pp.1337-1348.
- [9] Mohamed Magdy Mosbah, "Current Services in Cloud Computing: A Survey," *International Journal of Computer Science, Engineering and Information Technology (IJCEIT)*, Vol.3, No.5, October 2013.
- [10] Yang, Jianfeng, and Zhibin Chen, "Cloud computing research and security issues," Computational intelligence and software engineering (CiSE), 2010 international conference on. IEEE, 2010.
- [11] Chuang, I-Hsun, et al., "An effective privacy protection scheme for cloud computing," Advanced Communication Technology (ICACT), 2011 13th International Conference on. IEEE, 2011.
- [12] Azhad, Syed, and Mr Srinivas Rao, "Ensuring Data Storage Security in Cloud Computing," In Proceedings of National Conference on Computing Concepts in Current Trends, India, pp. 310-313. 2011.
- [13] Jain, A., Mishra, M. K., Peddoju, S. K., & Jain, N, "Energy efficient computing-green cloud computing," In Energy Efficient Technologies for Sustainability (ICEETS), IEEE, 2013, pp. 978-982.
- [14] Yeanf-Fu Wen, "On Energy Efficiency Data Access and Backup for Cloud Computing Networks," Green Computing and Communications (GreenCom), 2013 IEEE and Internet of Things (iThings/CPSCoM), IEEE International Conference on and IEEE Cyber, Physical and Social Computing, 20-23 Aug. 2013, pp. 1369 – 1374.
- [15] Riddhi Patel, Hitul Patel, Sanjay Patel, "Quality of Service Based Efficient Resource Allocation in Cloud Computing," *International Journal for Technological Research in Engineering*, vol. 2, issue 9, 2015.
- [16] Ashkan Paya and Dan C. Marinescu, "Energy-aware Load Balancing and Application Scaling for the Cloud Ecosystem," in Cloud Computing, IEEE Transactions on 2015, vol 99, pp 1-1.
- [17] F. Satoh, H. Yanagisawa, H. Takahashi and T. Kushida, "Total Energy Management system for Cloud Computing," Proceedings of the IEEE International Conference of the Cloud Engineering (IC2E), 2013, March 25-27; Redwood City, CA.
- [18] S. Srikantaiah, A. Kansal, and F. Zhao, "Energy aware consolidation for cloud computing," *Cluster Computing*, vol. 12, pp. 1–15, 2009.

- [19] Anton Beloglazov et al, "*Energy Efficient Allocation of Virtual Machines in Cloud Data Centers*", 2010 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing.
- [20] Shailesh S. Deore et al, "Energy-Efficient Job Scheduling and Allocation Scheme for Virtual Machines in Private Clouds," *International Journal of Applied Information Systems (IJ AIS)* – ISSN: 2249-0868 Foundation of Computer Science FCS, New York, USA vol 5, no.1, January 2013.
- [21] Ma, Wenqing, and Jing Zhang, "*The survey and research on application of cloud computing*," In Computer Science & Education (ICCSE), 2012, 7th International Conference on, pp. 203-206. IEEE, 2012.
- [22] Hengliang, Shi, Bai Guangyi, Liu Zhonghua, and Tang Zhenmin, "*Complex task query based on cloud computing Resource: Improved approximate skyline algorithm*," In Information Science and Engineering (ICISE), 2010 2nd International Conference on, pp. 1093-1096. IEEE.