Optimization of Tasks in Mobile Cloud Computing

Neha Gupta
M.Tech Scholar
Department of Information
Technology
Chandigarh Engineering
College
Mohali, Punjab, India

Parminder Singh
Assistant Professor
Department of Information
Technology
Chandigarh Engineering
College
Mohali, Punjab, India

Manveen Kaur
Assistant Professor
Department of Computer
Science and Technology
Shri Guru Teg Bhadur Khalsa
College
Anandpur Sahib, Punjab, India

ABSTRACT
Mobile cloud computing is a new platform for the execution of mobile applications where cloud performs the stimulating computing-intensive tasks and storing data in place of the mobile devices and provides them an illusion of infinite computing resources. This research work considers a cloud based mobile computing system consisting of virtual machines (as resources), cloudlets (as requests) and broker, where the broker assigns user cloudlet requests to virtual machines to be processed by the servers. It has been a great challenge to design and build an effective load balancing algorithm for the broker which spreads the service request load on virtual machines while utilizing the resources to the maximum. In this paper, a scheduling model to optimize the load based on maximum resource utilization rate using genetic algorithm for scheduling requests is proposed where the computing capacity of a datacenter is divided into number of tiers. Simulation results shows that the proposed work can effectively cope with the load imbalance problem in mobile cloud computing.

Keywords
Load Balancing, Cloudlets, Virtual Machines, Response Time.

1. INTRODUCTION
Recently the cloud computing technology has emerged as a new information technology infrastructure for the fast developing IT industry. In cloud computing, information is permanently stored in large-scale data centers on the Internet all over the world and temporarily accessed and cached on clients including desktops and portable PCs, sensors, etc [6]. Cloud computing distributes extremely scalable data, software, and hardware capabilities as a service to external clients with Internet accesses. And the highly scalable computation capability of the cloud data centers can further assist and accelerate most of our computation intensive services and works effectively. Therefore the cloud computing has been seen as the key technology to achieve economies of scale in the deployment and operation of IT solutions.

Recent advances in mobile communication technologies led a new wave of user demands for rich mobile service experience. Mobile users always expect broadband Internet access all the time, interact with each other via social networks while moving; furthermore, they are seeking access to a wealth of media-based contents and services. Furthermore the developments of innovative mobile services, e.g., rich media broadcasting, surveillance, mobile video streaming, gaming, e-health care, etc., can be greatly facilitated by mobile cloud computing platforms. Mobile devices because they are resource restricted naturally, it is essential for the cloud to provide computational support for many media-rich applications. The combination of mobile media and cloud computing highly arises many technical issues, and one of the major challenges is the problem of load optimization [15]. When a number of requests focused only on some dedicated severs while the others remain idle, the processing capability and the processing speed of a server both get affected leading to low response time of server and sometimes even its breakdown. In this paper a technique is introduced to evenly distribute the requests across the whole network and to speed up the processing by dividing the cloud datacenters computing capacity into various numbers of virtual machines and providing an effective allocating policy for mapping of requests with desired virtual machines thereby optimizing the load and reducing the response time and processing the requests all at a time with a focus on increasing the utilization of resources. The main area of interest of this paper is to devise an effective algorithm for such allocating scheme.

1.1 Terms and Concept used in Proposed Algorithm [13]:
Datacenter component is used for managing user or service requests (cloudlets).VM consist of application elements which are then connected with the requests, and allocated for VM process sharing by the datacenter's host components.

i) Datacenter
The data center is set of several computing servers that provide the resources. Cloud provider provides its computing services through its datacenter and datacenter houses servers or physical server or hosts. The main and important in datacenter is virtual machine which plays a very important role. Physical server deal with number of hosts and hosts in turn manage virtual machines. VM instance have been instantiated on the host and provides the resources to the consumers as per requirements.

ii) Host:
They can involve in the mechanism of organizing adequate memory and bandwidth to the process elements to process or execute them inside VM. Host is also accountable for construction and destruction of VMs.

iii) Virtual Machine (VM):
VM characteristics include memory, processor, storage, and VM scheduling policy. Multiple VM can run on single hosts concurrently and maintain processor allocation policies.

iv) Cloudlet
Cloudlets are used to deliver the data in cloud service model. Cloudlet is basically a request for any application from consumer through internet. Broker or hypervisor provide a unique id for a cloudlet and according to requirements allocate the resources and execute the cloudlet.

v) Broker
Broker acts as the intermediary module that provides the services among the consumers and provider. The broker works with consumers to help and understand the budgeting

of resources, requirement of data management, services, process of work etc. Broker helps the consumers to provide the list of cloud providers. Broker manages the resources, allocating and reallocating the resources (virtual machine).

2. RELATED WORK

Scheduling of tasks to balance the load is a critical issue in Mobile Cloud Computing, and has gained a lot of attention in recent years. A range of researches have been done in this region. Even though a plethora of works available in the literature, a handful of important research works are reviewed here.

Xiaocheng Lui et al. [17] have implemented Conservative Migration and Consolidation supported Backfilling CMCBF to increase the utilization of resources in the data-centers when two or more jobs are running in parallel. Meanwhile the algorithm tries to use the computing capacity of the data center nodes running parallel processes with low resource utilization to improve the performance of the job scheduling. The algorithm uses virtualization technologies to divide the computing capacity of each node into two levels i.e foreground virtual machine and the background virtual machine. It ensures that a job is dispatched to run in foreground VMs whenever the number of required VMs is free or occupied by jobs arriving later than it satisfies its node requirement or they can also dispatched to background VMs to run if the VM requirement satisfies or else their state are suspended to run later when the required number of VMs are free. It also allows jobs to run in background VMs simultaneously with those foreground VMs to improve node utilization.

Brototi Mondala et al. [7] proposed a soft computing based load balancing optimization approach Known as Stochastic Hill climbing that allocates incoming jobs to the servers or virtual machines (VMs).A stochastic and Local Optimization algorithm is simply a loop that constantly moves in the way of growing value. It stops when it reaches a peak where no neighbor has a higher value. Each element of the set is evaluated according to some criteria designed to move closer to a valid assignment. The best element of the set is made the next assignment. This basic operation is repeated until either a solution is found or a stopping criterion is reached. So it has two main components a candidate generator which maps one solution candidate to a set of possible successors, and a evaluation criteria which ranks each valid solution, such that improving the evaluation leads to better solutions

Ram Kumar et al. [16] proposed Effective Resource Utilization Algorithm (ERUA). This algorithm based upon the 3 tier architecture of cloud. In three tier consumers, resources providers and service providers. The cloud users rely on the quality of service and providers rely on profits. To accomplish these, need an efficient algorithm which reduces the waiting time, cost and maximizes the utilization of resources effectively. In this algorithm when many users submit their request to a provider, then it passes to classifiers that divides into many tasks and assigned a random priority. After that, it scheduled the tasks to the scheduler. Scheduler has many scheduling units and each having their own priority high, low and medium. Scheduler assigns the tasks to the schedule units and execute according to priority. The task of lesser limit will be scheduled first. After that compactor collects the tasks and send it to the provider.

3. SYSTEM MODEL

In this system model cloud provides services through its data centers which consist of a collection of number of computing servers as virtual machines that basically provides resources to the users to process their requests. When a request comes to a cloud it is then managed by the broker, the broker submits the requests to the appropriate virtual machine according to the scheduling scheme of the cloud as shown in figure 1.

4. SEQUENCE OF OPERATIONS

1. User interacts to a Cloud via Broker. User asks for VMs to fulfill its resource requirements. User ask broker for VMs (resources). Broker then maps the request to the virtual machine according to the scheduling Algorithm.

2. Two array lists are created one is for containing the list of virtual machines over which tasks are to be scheduled to execute and other is for containing the list of requests to be processed.

3. N tasks are divided into M classes, the special kind of tasks are put together and make up a set, the tasks of large amount of calculation are put together and make up a set and the tasks of small amount of calculation are put together and makeup a set, it has M sets.

4. M tasks are scheduled to virtual machines at the same time, the tasks of large amount of calculation are scheduled to resources queue whose calculation ability are strong., the tasks of small amount of calculation are scheduled to resources queue whose calculation ability are weak.
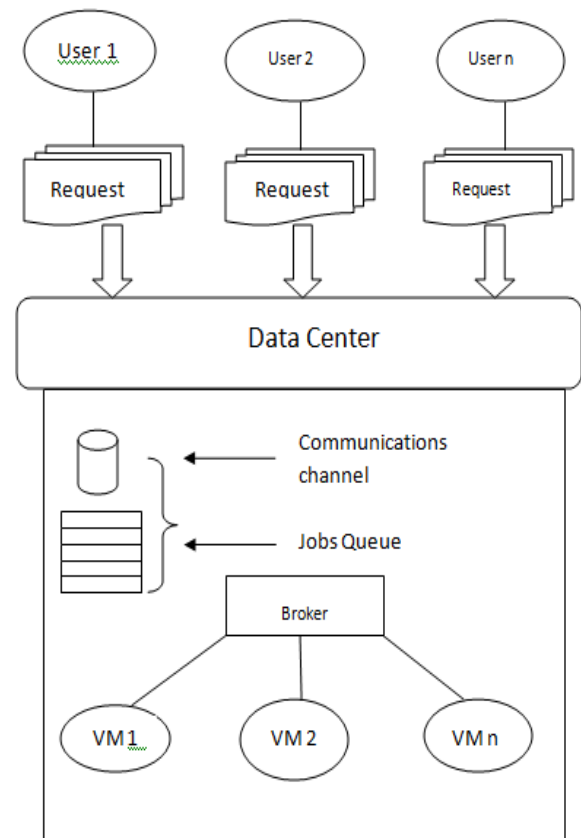

Figure 1: System Model

5. SIMULATION

Assumptions:

a. We consider only one datacenter per cloud service provider, although it can be easily extended to include multiple datacenters.

b. For simplicity, we assume that the physical servers in datacenters are of similar pattern and have the capability of running the same number of VM instances.

c. The resource requests are in terms of VM instances, although cloud users consume services ranging from Infrastructure-as-a-Service (in terms of physical server instances) to Software-as-a-Service (higher order applications and services). We assume that the "Broker" will translate such requests into VM instances required to provide those services.

The following parameters were considered during simulation:

• Number of cloud service providers: 1

• Number of physical servers per datacenter: 10~50

• Maximum Virtual Machines per server: 5

• Resource request quantum: 10~50 vms per request

• Resource request frequency: 2~5 per minute

• Duration of resource usage: 30~60 minutes

• Flash-crowd scenario frequency: once every 3 hours

• Flash-crowd scenario duration: 10 minutes

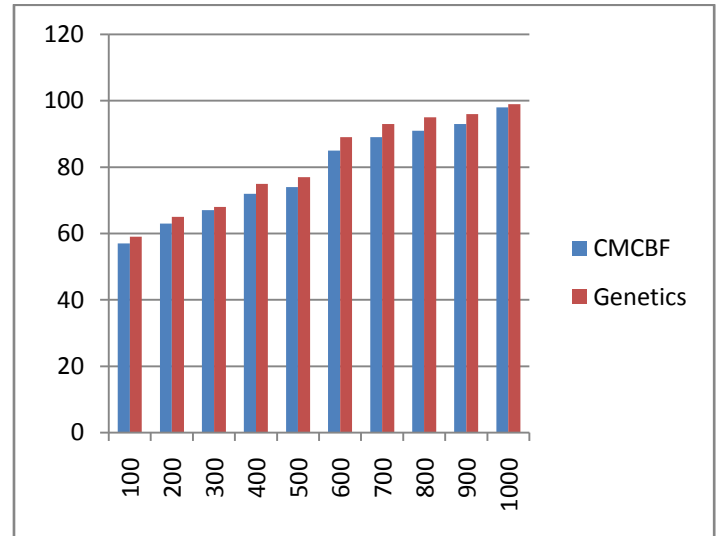• Flash-crowd resource request frequency: 15~20 per minutes

6. RESULTS

This section gives the performance analysis of the proposed approach based on the task optimized scheduling through genetic algorithm. To understand the significance of the proposed approach it has to be analyzed with an existing method. Here, we choose the load balancing scheduling method CMCBF to evaluate the significance of the proposed approach. The comparative analysis is subjected over the node utilization rate of the proposed approach and existing approach. Node utilization is the average vm utilization, which is the fraction of busy CPU cycles in the simulation. The analyses are conducted over the virtual machines in the datacenter of the simulated cloud network. The node utilization rate is plotted in the figure 2. The analysis shows that, the proposed approach utilizes the node utilization rate more efficiently than the existing approach. So, considering load balanced condition and the node utility rate the proposed approach has better efficiency over the existing approach.

Table 1: Values of resource utilization for CMCBF and Genetic Algorithm

Cloudlets Processed (In numbers)	CMCBF (Resource utilization %)	Proposed (Resource utilization %)
100	57	59
200	63	65
300	67	68
400	72	75
500	74	77
600	85	89
700	89	93
800	91	95
900	93	96
1000	98	99

Figure 2: Comparison graph showing comparison between CMCBF and Genetic Algorithm in terms of resource utilization rate

7. CONCLUSION

The proposed scheme implemented to schedule the requests on different resources while optimizing the load of the system in the mobile cloud computing system to provide access to resources to the clients. But the existing schemes suffers from various overhead and fails when there are number of requests competing at a time and the resources available are not sufficient to service the requests. The proposed approach overcomes the said limitations. The performance of the proposed approach will be evaluated in terms of the node utilization. Higher the utilization means more number of requests can be processed by the server. Simulation results show the effectiveness of the proposed approach and an increase of 3% in the utilization of resources had been noticed. Therefore, the proposed work improves the performance of the service by achieving the mentioned objectives.

Algorithm	CMCBF	Genetic Algorithm
Resource Utilization %	78.9%	81.6%

8. REFRENCES

[1] Amandeep Kaur, "Analysis of Load Balancing Techniques in Cloud Computing", International Journal of Computers & Technology, vol.4, no.2, pp.2277-3061, March-April, 2013.

[2] Arash Ghorbannia Delavar, Mahdi Javanmard , Mehrdad Barzegar Shabestari and Marjan Khosravi Talebi "Reliable Scheduling Distributed In Cloud Computing" International Journal of Computer

Science Engineering and Applications (IJCSEA) vol.2, no.3, pp.1-16, June 2012.

[3] B. Subramani, "A New Approach For Load Balancing In Cloud Computing", IEEE vol.2, pp.1636-16405, May 2013.

[4] Dihal, Soebhaash, "Mobile cloud computing: state of the art and outlook", Emerald Group Publishing Ltd, vol.15, pp. 4-16, 2013.

[5] Dr. M. Dakshayini, Dr. H. S. Guruprasad "An Optimal Model for Priority based Service Scheduling Policy for Cloud Computing Environment" International Journal of Computer Applications, vol.32, no.9, pp.23-29, October 2011.

[6] El-Sayed T, El-kenawy, Ali Ibraheem, El-Desoky, Mohamed F and Al-rahamawy, "Extended Max-Min Scheduling Using Petri Net and Load Balancing" International Journal of Soft Computing and Engineering (IJSCE), vol.2, September 2012.

[7] Entezari-Maleki and Saeed Parsa, "RASA: A New Task Scheduling Algorithm in Grid Environment" World Applied Sciences Journal, vol.2, no.7, pp.152-160, 2009.

[8] Gaochao Xu, Junjie Pang and Xiaodong Fu," A Load Balancing Model Based on Cloud Partitioning for the Public Cloud" Tsinghua Science And Technology, vol.18, no.2, pp.34-39, February 2013.

[9] Gaurav Raj, Ankit Nischal, "Efficient Resource Allocation in Resource provisioning policies over Resource Cloud Communication Paradigm" International Journal on Cloud Computing: Services and Architecture (IJCCSA), vol.2, no.3, pp.11-18, June 2012.

[10] Huber Flores, Satish Narayana Srirama, Carlos Paniagua, "Towards mobile cloud applications: Offloading resource intensive tasks to hybrid clouds", International Journal of Pervasive Computing and Communications, vol. 8, pp. 344-367, 2012.

[11] Ioannis Moschakis, A.Karatza, D.Helen, "Performance and Cost evaluation of Gang Scheduling in a Cloud Computing System with Job Migrations and Starvation Handling" IEEE, vol.4, pp.418-423, June-July 2011.

[12] Kahina Bessai, Samir Youcef, Ammar Oulamara, Claude Godart and Selmin Nurcan, "Bi-criteria workflow tasks allocation and scheduling in Cloud computing environments", IEEE, vol.4, pp.638-645, 2012.

[13] Kyi, Mon Hsu and Naing Thinn Thu, "Stochastic Markov Model Approach For Efficient Virtual Machines Scheduling On Private Cloud", International Journal on Cloud Computing: Services and Architecture(IJCCSA), vol.1, no.3, pp.1-13, November 2011.

[14] Linan Zhu, Qingshui Li and Lingna, "Study on Cloud Computing Resource Scheduling Strategy Based on the Ant Colony Optimization Algorithm" International Journal of Computer Science Issues (IJCSI), vol. 9, no.2, pp.1694-0814, September 2012.

[15] Luiz F.Bittencourt, Edmundo R.M.Madeira and Nelson L.S da Fonseaca, "Scheduling in Hybrid Clouds" IEEE, pp.42-47, September,2012.

[16] Ramkumar.N, Nivethitha.S, "Efficient Resource Utilization Algorithm (ERUA) for Service Request Scheduling in Cloud" International Journal of Engineering and Technology (IJET), pp.1321-1327, May 2013.

[17] Xiaocheng, Lui, Cheng, Wang, Juliang Chen and Albert Y.Zomaya, "Priority Based Consolidation of Parallel Workloads in the Cloud" IEEE, vol.24, no.9, pp.1874-1882, September 2013.