International Journal For Technological Research In Engineering
Volume 2, Issue 1, September-2014 *ISSN (Online): 2347 - 4718*

A REVIEW ON SCHEDULING TECHNIQUES AND RESOURCE ALLOCATION STRATEGIES IN MOBILE CLOUD COMPUTING SCHEDULING TECHNIQUES & RESOURCE ALLOCATION STRATEGIES IN MCC

Neha Kapoor[1], Parveen Kakkar[2]
[1]Research Scholar, [2]Asst. Prof.,CSE
DAV Institute of Engineering & Technology, Jalandhar, India

Abstract: Mobile Cloud computing has evolved as a new age technology that has influenced the enterprises and markets in a profitable way. The data and the applications can be accessed from anywhere with the help of cloud technology. Companies rent resources from cloud for storage and other computational purposes based on pay-as-you-go model. Mobile cloud computing is also based on the basic cloud computing concepts where all the data processing and storage happens in the cloud, outside the mobile device having scarcity of resources. The major issues in mobile cloud computing are that which scheduling technique must be adopted and how the resources can be efficiently allocated thus minimizing the cost incurred. In this paper, the various scheduling techniques and resource allocation strategies are discussed.
Keywords: Mobile Cloud Computing, Cloud Services, Scheduling Techniques, Resource Allocation.

I. INTRODUCTION

Cloud computing has emerged as a computing paradigm which aims to provide reliable, customized and QoS (Quality of Service) guaranteed computing dynamic environments for end-users [2]. Cloud computing can be defined as the aggregation of computing as a utility and software as a service [5] where the applications are delivered as services over the Internet and the hardware and systems software in data centers provide those services [6]. Also called 'on demand computing', 'utility computing' or 'pay as you go computing', the concept behind cloud computing is to offload computation to remote resource providers. Mobile cloud computing is also based under the basic cloud computing concepts where all the data processing and storage happens in the cloud, outside the mobile device having scarcity of resources. As discussed in [10] there are certain requirements that need to be met in a cloud such as adaptability, scalability, availability and self-awareness.

II. CLOUD COMPUTING SERVICES

Generally, a cloud computing is a large-scale distributed network system implemented based on a number of servers in data centers. The cloud services are generally classified based on a layer concept (Fig. 1).

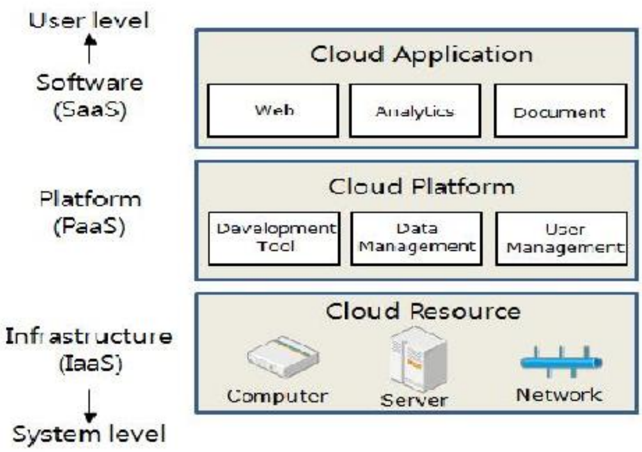


Fig 1: Cloud Computing Services [13]

Sung-Min Jung et al [13] in their paper have explained the various layers of the cloud paradigm; Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) are stacked.

A. Infrastructure as a Service (IaaS): IaaS enables the provision of storage, hardware, servers and networking components.[13] The client typically pays on a per-use basis. Thus, clients can save cost as the payment is only based on how much resource they really use. Infrastructure can be expanded or shrunk dynamically as needed. The examples of IaaS are Amazon EC2 (Elastic Cloud Computing) and S3 (Simple Storage Service).

B. Platform as a Service (PaaS): PaaS offers an advanced integrated environment for building, testing and deploying custom applications.[13] The examples of PaaS are Google App Engine, Microsoft Azure and Amazon Map Reduce/Simple Storage Service.

C. Software as a Service (SaaS): SaaS supports a software distribution with specific requirements. In this layer, the users can access an application and information remotely via the Internet and pay only for that they use. [13] Sales force is one of the pioneers in providing this service model. Microsoft's Live Mesh also allows sharing files and folders across multiple devices simultaneously.

III. MOBILE CLOUD COMPUTING ARCHITECTURE

N. Fernando et al. [11] had proposed the architecture of Mobile Cloud Computing for an operational mobile cloud, which exploits the locally available mobile resources, while ensuring user privacy, security, and operates at optimum cost; their proposed system is mainly composed of five components: Job handler, resource handler, cost manager, privacy and security manager, and context manager.

A. Resource handler: It shall be responsible for searching for and discovering other mobile resources, connecting, maintaining connections and communicating with the external mobile devices, and also monitoring them for potential nodes entering or leaving the cloud.
B. Cost manager: It must determine the user priorities (e.g.:battery conservation, fast execution, monetary gain) and by taking into account the job at hand, available resources, and required resources, come to a decision whether to offload or not.
C. Job handler: It dynamically partitions the application and or data set required, offloads the generated jobs, and maintains the job pool. Of course, for these modules to function, intercommunication is needed.
D. Privacy and security manager: It needs inputs from sensors and user in addition to network interfaces to determine the most suitable policy to enforce.
E. Context Manager: In order to ensure cost efficiency the context manger manages the context and system resources according to the user's requirements.

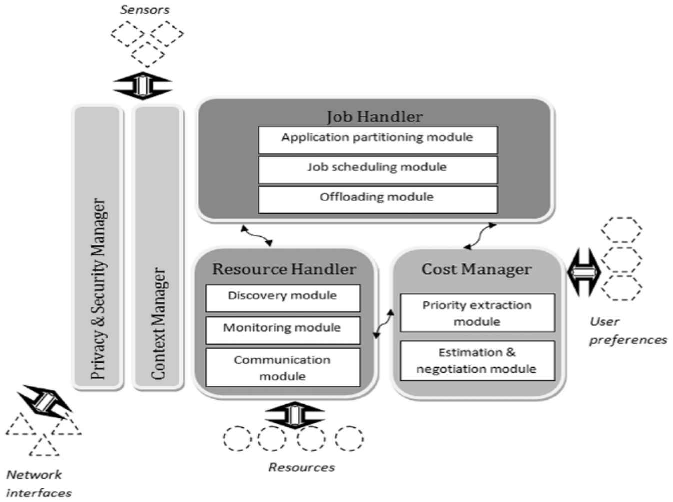


Fig 2: Mobile Cloud Computing Architecture [11]

IV. SCHEDULING TECHNIQUES

Scheduling techniques defined as the decision that should be taken by Scheduling Manager when the assigned node leaves the cloud. Naif Aljabri et al. [1] proposed two scheduling techniques. These techniques are:

A. Recovery Technique

In this technique, the Scheduling Manager assigns the jobs to the proper node based on the job size and the node rate value. While the job is running, the Node Tracker communicates with Scheduling Manager and sends the job stat periodically. The Scheduling Manager will save the stat temporary until the job is done. During the job execution, if the communication lost with the Node Tracker (node leave the cloud) , the Scheduling Manager assign the last job stat to a new idle node and continue the job execution from the last stat. "Fig. 3", shows the scenario for this technique and "Fig. 4", illustrate the activity diagram for this technique.

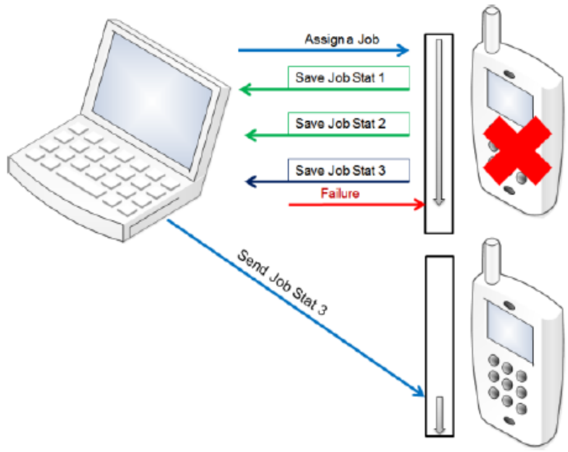


Fig 3: Recovery Scheduling Technique [1]

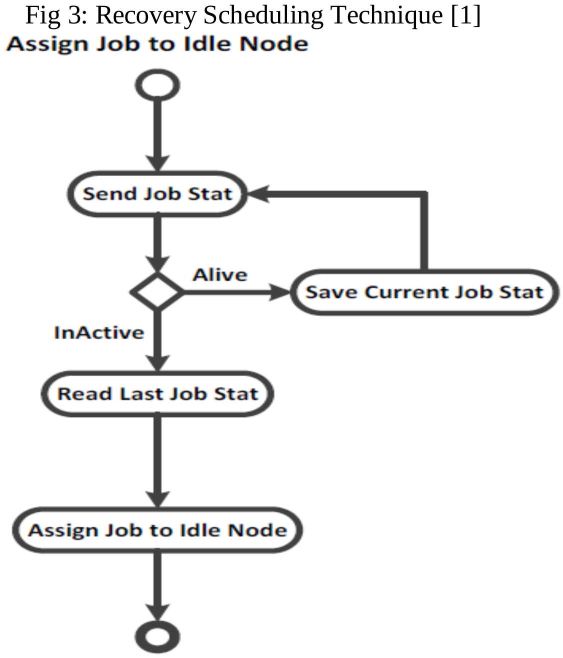


Fig 4: Activity Diagram for Recovery Technique [1]

B. Fault-tolerance Technique

In this technique, the Scheduling Manager use two nodes for each job. Every node run a copy of the same job. The Scheduling Manager use a periodic heartbeat call to the Node Tracker to show that the nodes are alive. if one of the nodes lost (node leave the cloud) , the Scheduling Manager read the current job stat from the active node and assign it to a new idle node to continue the job execution together with the active node. "Fig. 5", shows the scenario for this technique and "Fig. 6", illustrate the activity diagram for this technique.

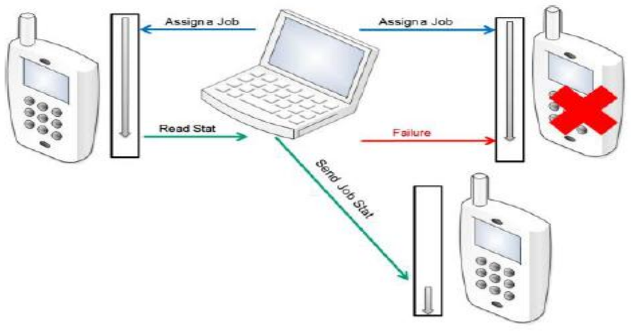
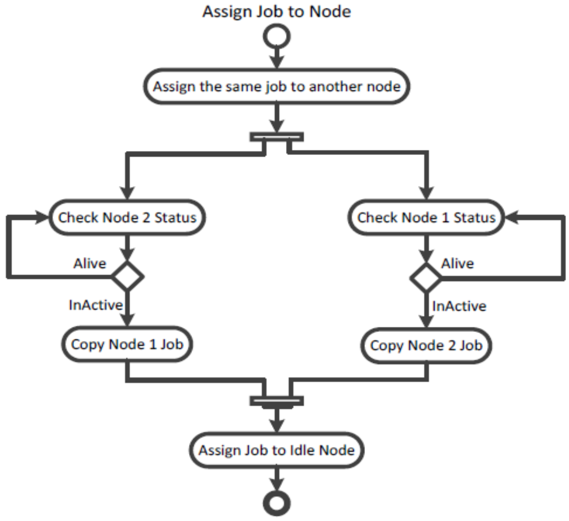


Fig 5: Fault Tolerance Technique [1]



.

Fig 6: Activity Diagram for Fault Tolerance Technique [1]

V. RESOURCE ALLOCATION STRATEGIES

V. Vinothina et al. [2] had discussed various resource allocation strategies used in mobile cloud computing. The input parameters to RAS and the way of resource allocation vary based on the services, infrastructure and the nature of applications which demand resources. The schematic diagram in Fig.7 depicts the classification of Resource Allocation Strategies (RAS) proposed in cloud paradigm. The following section discusses the RAS employed in cloud.

A. Execution Time

In the work by Jiani at.al [3], actual task execution time and preemptable scheduling is considered for resource allocation. It overcomes the problem of resource contention and increases resource utilization by using different modes of renting computing capacities. But estimating the execution time for a job is a hard task for a user and errors are made very often [4].

B. Policy

Since centralized user and resource management lacks in scalable management of users, resources and organization-level security policy [5], Dongwan et al. [5] has proposed a decentralized user and virtualized resource management for IaaS by adding a new layer called domain in between the user and the virtualized resources. Based on role based access control (RBAC), virtualized resources are allocated to users through domain layer.

C. Virtual Machine

A system which can automatically scale its infrastructure resources is designed in [6]. The system composed of a virtual network of virtual machines capable of live migration across multi- domain physical infrastructure. By using dynamic availability of infrastructure resources and dynamic application demand, a virtual computation environment is able to automatically relocate itself across the infrastructure and scale its resources.

D. Gossip

Cloud environment differs in terms of clusters, servers, nodes, their locality reference and capacity. The problem of resource management for a large-scale cloud environment (ranging to above 100,000 servers) is addressed in [7] and general Gossip protocol is proposed for fair allocation of CPU resources to clients.

E. Utility Function

There are many proposals that dynamically manage VMs in IaaS by optimizing some objective function such as minimizing cost function, cost performance function and meeting QoS objectives. The objective function is defined as Utility property which is selected based on measures of response time, number of QoS, targets met and profit etc.

F. Hardware Resource Dependency

In paper [8], to improve the hardware utilization, Multiple Job Optimization (MJO) scheduler is proposed. Jobs could be classified by hardware-resource dependency such as CPU-bound, Network I/O-bound, Disk I/O bound and memory bound. MJO scheduler can detect the type of jobs and parallel jobs of different categories. Based on the categories, resources are allocated. This system focuses only on CPU and I/O resource. Eucalyptus, Open Nebula and Nimbus are typical open source frame works for resource virtualization management [9]. The common feature of these frameworks is to allocate virtual resources based on the available physical resources, expecting to form a virtualization resource pool decoupled with physical infrastructure.

G. Auction

Cloud resource allocation by auction mechanism is addressed by Wei-Yu Lin et al. in [10]. The proposed mechanism is based on sealed-bid auction. The cloud service provider collects all the users' bids and determines the price. The resource is distributed to the first kth highest bidders under the price of the (k+1)th highest bid.

H. Application

In the work by Tram et al. [11], Virtual infrastructure allocation strategies are designed for workflow based applications where resources are allocated based on the workflow representation of the application. For work flow based applications, the application logic can be interpreted and exploited to produce an execution schedule estimate. This helps the user to estimate the exact amount of resources that will be consumed for each run of the application. Four strategies such as Naive, FIFO, Optimized and services group optimization are designed to allocate resources and schedule computing tasks.

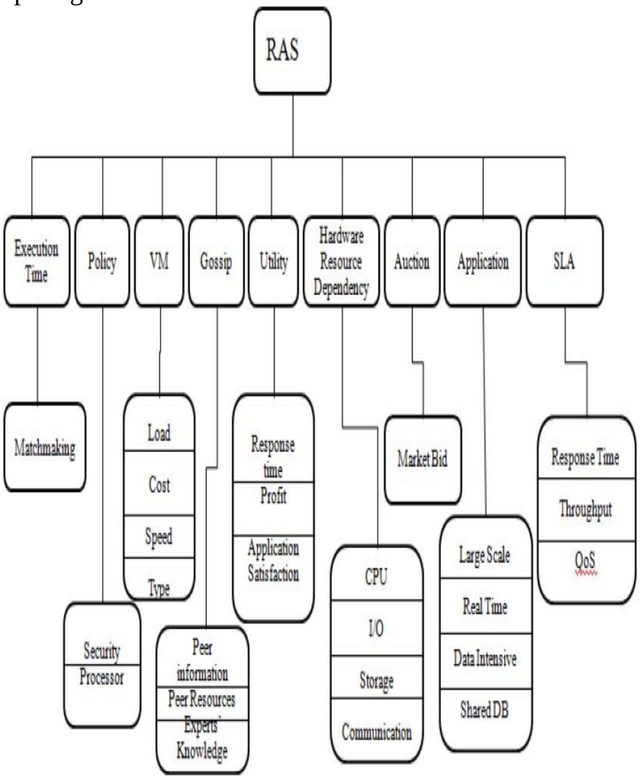


Fig 7: Resource Allocation Strategies [2]

VI. ISSUES RELATED TO MCC

A. Operational Issues

Operational issues refer to underlying technological matters such as the method of offloading computations, cost–benefit models that aid in taking the decision to offload or not, how the mobility of devices is managed/supported, and connection protocols used.

B. End User Issues

End users issues relate to issues that directly involve users such as incentives for participating, interoperability and cost. When using a mobile cloud, one of the key challenges experienced by the end users is the transaction infrastructure.

C. Service and Application Level Issues

Service and application level issues relate to the factors concerned with performance measurements of the system, and the QoS of the system.

D. Context Awareness

Schilit et al. [12] describes the three important aspects of context as: the user's location, other users in the vicinity, and the resources in the user's environment.

E. Data Management Issues

For many cloud users and providers, managing data on the cloud raises many complications. In mobile cloud computing, as the name itself suggests, data that would traditionally be only accessible only to the mobile device's owner, would now be stored on, accessible to, shared with external devices or users.

VII. CONCLUSION

In mobile cloud paradigm, there is a need for an effective resource allocation strategy which is required for achieving maximum profit for cloud service providers and to maximize the user satisfaction. This paper summarizes the architecture, scheduling techniques, Resource Allocation Strategies and issues of mobile cloud computing.

REFERENCES

[1] Naif Aljabri et al., "Scheduling Manager for Mobile Cloud Using Multi-Agents", International Journal of Computer and Information Technology (ISSN: 2279 – 0764) Volume 02– Issue 03, May 2013.
[2] V. Vinothina et al. "A Survey on Resource Allocation Strategies in Cloud Computing", International Journal of Advanced Computer Science and Applications, Vol. 3, No.6, 2012.
[3] Jiyani et al. "Adaptive resource allocation for preempt able jobs in cloud systems", (IEEE, 2010), pp.31-36.
[4] Shikharesh Majumdar "Resource Management on cloud: Handling uncertainties in Parameters and Policies" (CSI communicatons, 2011, edn, pp.16-19.
[5] M. Satyanarayanan, P. Bahl, R. Caceres, N. Davies, The case for VM-based cloudlets in mobile computing, IEEE Pervasive Computing 8 (2009) 14–23.
[6] Dongwan Shin and HakanAkkan "Domain- based virtualized resource management in cloud computing".
[7] P.Ruth,J.Rhee, D.Xu, R.Kennell and S.Goasguen, "Autonomic Adaptation of virtual computational environments in a multi-domain infrastructure", IEEE International conference on Autonomic Computing, 2006,pp.5-14.
[8] RerngvitYanggratoke, FetahiWuhib and Rolf Stadler: "Gossip-based resource allocation for green computing in Large Clouds:" 7th International conference on network and service management, Paris, France, 24-28 October, 2011.
[9] Weisong Hu et al. " Multiple Job Optimization in MapReduce for Heterogeneous Workloads " IEEE Sixth International Conference on Semantics, Knowledge and Grids 2010,pp.135-140.
[10] XiaoyiLu,Jian Lin, Li Zha and ZhiweiXu: Vega Ling Cloud, "A Resource Single Leasing Point System to Support Heterogenous Application Modes on Shared Infrastructure", (IEEE,2011),pp.99-106.

[11] Niroshinie Fernando et al. , "Mobile cloud computing: A survey", Future Generation Computer Systems 29 (2013) 84–106.
[12] B. Schilit, N. Adams, R. Want, "Context-aware computing applications", Proceedings of the Workshop on Mobile Computing Systems and Applications, IEEE, 1994, pp. 85–90.
[13] Sung-Min Jung et al., "Applying Scheduling Algorithms with QoS in the Cloud Computing", 978-1-4799-0604-8/13, IEEE 2013.

