Resource Provisioning Issues in Mobile Cloud Computing: A Survey

Merugu.Gopichand
Professor &Head, Department of IT,VardhamanCollege Engineering , SamshabadTelagana,INDIA,
gopi_merugu@yahoo.com

Abstract:

Resource provisioning is a critical concern for mobile users as they have less capability in providing high battery, bandwidth and memory. Mobile cloud computing provides a solution to this problem as it brings all cloud services to mobile users and mobile service providers. This paper discusses different issues in resource provisioning so as to manage the usage of applications effectively by mobile users. It also highlights research trends and challenges in managing the resources for mobile users in cloud.

Key Words: Resource Provisioning, Mobile cloud Computing,Resource Management

1 Introduction

A simple way to define "Mobile Cloud computing is "A Mobile cloud computing is a structure where both the data storage and the data processing happen outside the mobile device. Mobile cloud applications move the computing power and data storage away from mobile phones to the cloud". Aepona [1] describes Mobile Cloud Computing as a new paradigm for mobile applications where the data processing and storage are moved from the mobile device to powerful and centralized computing platforms located in clouds. These centralized applications are then accessed over the wireless connection based on a thin native client or web browser on the mobile devices. WeiweiJia [2] defines that, the benefits brought by Cloud Computing have been also demonstrated by the emergence of Mobile Cloud Computing, which is regarded as one of most disruptive technology for future mobile applications. Different from the general cloud computing concept, mobile cloud computing refers to an emerging infrastructure where both data storage and data processing happen outside of the mobile device from which an application is launched. Alternatively, Jacson H. Christensen [3] and L. Liu, R. Moulic, and D. Shea [4] Mobile Cloud Computing can be defined as a combination of mobile web and cloud computing which is the most popular tool for mobile users to access applications and services in the internet.

Mobile Cloud Computing provides the features of cloud computing to both mobile users and mobile service providers. It makes the computational process of mobile users very easy and manages the data in cloud. The resource provisioning is a process of discovering, allocation and monitoring resources like CPU, Memory, Storage etc in cloud and it is major challenge for mobile users who need their services to be carried out uninterruptedly. The mobile user's faces different problems of having less battery power, offloading of data because of less memory backup and less processing speed which makes them to depend on the cloud resources. The cloud computing considers the resource provisioning by considering different QoS factors[5] such as availability,throughput,security,reponse time, reliability and performance.

The Fig 1 gives the Mobile Cloud Computing architecture where in the mobile users access the cloud services directly from their applications or through mobile service provider using their mobile data. Cloud provides it major services like SaaS (Software as a Service), PaaS (Platform as a Service) and IaaS (Infrastructure as a Service) in collaboration with RaaS (Resource as a Service) [6] which deals with many issues how to provision resources, pricing schemes and migration of resources for effective usage of cloud for mobile users.

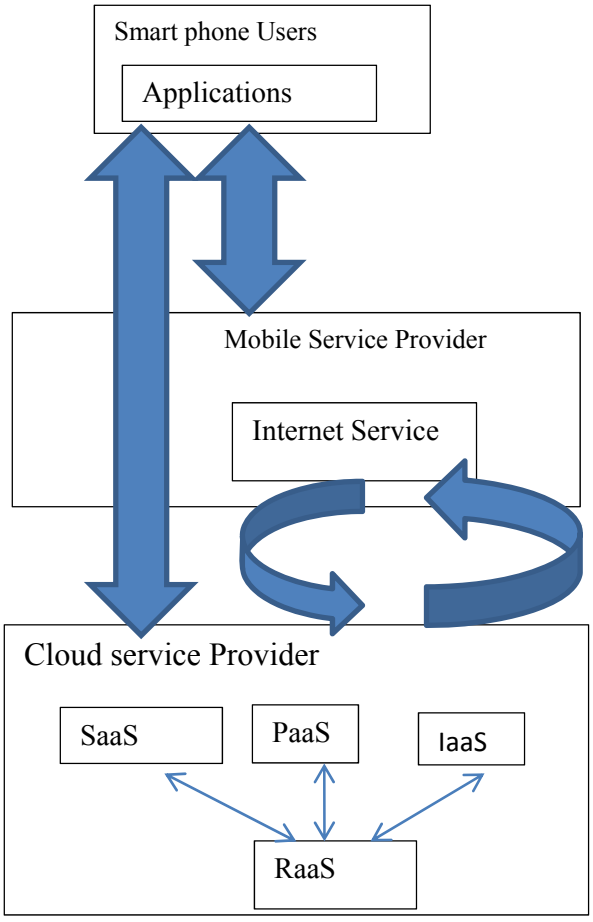

Fig 1.Mobile Cloud Computing Architecture

2 Resource Provisioning issues in MCC

The popular issues which need to be dealt in MCC for provisioning resources to mobile users are resource demand prediction, workload monitoring, workload Schedule, Resource Migration and Resource Optimization. Each of this is a challenge or a research issue which need to be resolved to mobile users or providers at the earlier stages of using resources provided by cloud providers.

2.1 Resource Demand Prediction

The Mobile users browsers many of the applications like Facebook, twitter, you tube etc. without any interruption on their mobile device. To allow uninterrupted supply of services to mobile users the mobile service provider and cloud service provider should collaboratively predict the resources needed for mobile clients and allocate resources like CPU, memory or I/O devices. The CRAM algorithm [7] is a greedy heuristic approach that generates a near-optimal solution for predicting the resources using simulated annealing technique. This algorithm is tested on two categories of applications like intensive computing and streaming.

To allocate resources dynamically and effectively to mobile users there is a need of accurate resource prediction. A self -adaptive prediction method using ensemble model and subtractive-fuzzy clustering based fuzzy neural network [8] is used to analyze the characters of user preferences and demands and then allocate resources. In this method a predictive model is constructed and a learning algorithm of fuzzy neural network is developed for effective prediction of resource demands.

2.2 Work Load Monitoring

It is important task of cloud service provider for monitoring workload requirements of mobile users after allocating resources. In agent based approach [9] the cloud directory service is used to update the database of virtual machine instances available for use in cloud. Virtual machine instances in the cloud provide a run time environment for agent based application partitions. They a provide a platform as a service(PaaS) ,rather than Software as a service (SaaS),and the only requirement they need to satisfy is to provide an isolated container(such as a JVM) for each offloaded partition to execute in.

In game theoretic model [10] the resource monitoring is done by an admission control system which uses linear programming formulation for resource sharing and a distributed algorithm which chooses each mobile user as a player who is chosen randomly. To effectively monitor resources during computation offloading the monitoring should be done at both at mobile device and cloud. A computation offloading framework to optimize make span in mobile cloud computing environment [11] provides a augmented execution of mobile applications using cloud resources by application partitioning and monitoring resources both at mobile and cloud using a genetic algorithm to find optimum offloading scenario.

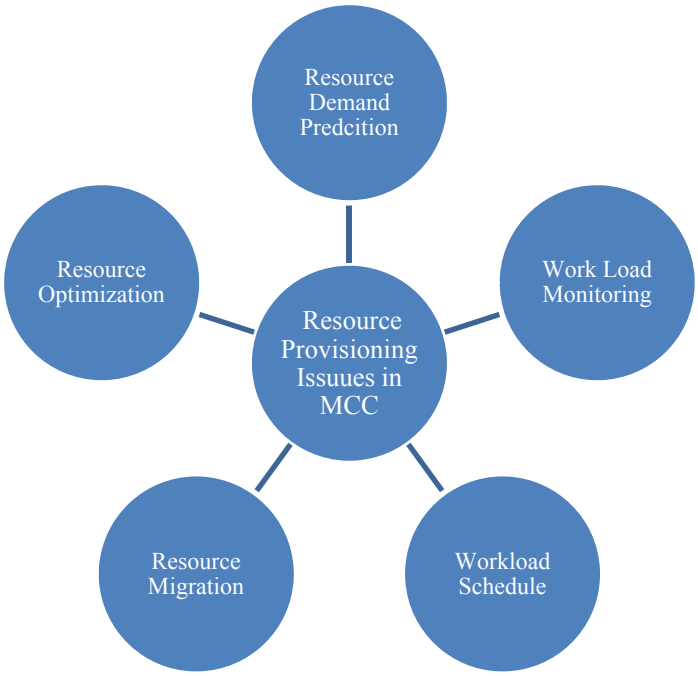

Fig 2:Resource Provisioning Issues in MCC

In Adaptive resource provisioning [12] the resource monitoring is based on semi-Markov decision process in which not only incomes of accepting services, but also the cost resulted from VM occupation in the cloud, other factors like processing time in cloud and mobile device battery consumption is also taken into account.

2.3 Work Load Schedule

Load schedule is a job of cloud service provider where in need to take in to consideration different mobile users simultaneously accessing application in cloud. Many algorithms were proposed to take care of workload schedule in both static and dynamic allocation of resources to mobile users in cloud.

In mobile cloud computing environment the load balancing is a critical task as it need to distribute workload evenly among all mobile users either statically or dynamically. A high level of user satisfaction and resource utilization can be achieved by ensuring an efficient fair allocation of computing resources. The appropriate load balancing helps in minimizing resource consumption, implementing fail-over, enabling scalability and avoiding bottlenecks. A Prognostic load balancing [13] is proposed by Ahmad et.al in which a simple exponential smoothing forecasting method is used to predict the workload requirements of mobile users and then schedule tasks for execution.

A cloud simulation scheduling algorithm [14] is based on multi-dimension QoS (Quality of Service) at first analytic hierarchy process in economic field was introduced into resource scheduling algorithm in order to

compute every dimensional parameters weight, then tasks was allocated to appropriate resource according to user satisfaction, QoS distance and loading equilibrium, etc.

In Adaptive resource provisioning the resource allocation is done based on semi-markov decision process to achieve the optimal policy for mobile Service providers. For a single request of mobile user it allows multiple cloud resources to get allocated based on their availability and maximize the resource utilization to enhance user experience. It also applies the economic way of achieving good rewards for the use of cloud resources by mobile users.

In Game Theoretic model the resource allocation based on biding and calculation of utility function for all mobile user requests. The utility function decides the allocation of cloud resources and gives the no of cloud resources assigned to the user. The mobile service providers reserve in advance the computing resources in cloud and all users need to undergo through admission control process for availability of computing resources and then get allocated based on the user requirements.

In agent based approach each mobile Service provider is treated as mobile agents who have capability for probing request for specific resource on particular cloud host. Once the mobile agent is capable of finding the resource it can clone itself on to multiple cloud hosts and do performance comparison for different hosts and determine the most promising host for specific application and allocate the resources.

2.4 Resource Migration

Resource Migration is an important task as cloud provider need to manage the over provision and under provision of resources. Whenever there is need of resources for mobile service providers to manage the huge traffic requests of mobile users they need to allocate or de-allocate depending on their requirements. The cloud service provider need to manage the resource migration between intra or inter cloud federation. Intra cloud resource migration is flexible for all users who are in the same domain or in intranet but inter cloud resource migration is difficult to manage as it need to cross boundary and move to another cloud and understand their provision and allocate resources.

Resource Migration provides better solution of managing fault tolerance issues occurring in cloud. To understand the management of fault tolerance models [15] a fault injection module for cloud Sim tool is tested and validated for simulated infrastructure. In this faults are introduced by event-driven model based on statistical distributions. This helps in understanding inter–cloud migration by intentionally inserting faults in resources of cloud and migrating to another resource to manage tasks.

Resource migration occurs not only based on mobile users request for resources but also based on pricing models charging for effective resources usage. This problem is resolved by managing multiple resource pools in a multi-tenant environment [16].Its benefit is reduction of resource waste by reusing already allocated resources available in the pool. The pricing model really benefits the cloud service providers in offering different pricing strategies that can have different effects on its revenue.

In MCC the resource migration is effectively managed by game theoretic approach [17] where in all resources tailored in to three tier layers like mobile client, Cloudlet middleware and Cloud servers. The Mobile clients get benefit offloading their computations from mobile nodes to external servers based on their time latency of taking their requests and allocating resources. This proposes a model to capture the user interactions and investigate the effects of computation offloading on user's perceived performance. Here, a usage scenario is considered where no central authority exists and multiple non cooperative mobile users share the limited resources of a close by cloudlet and selfishly decide to send their computations to any of three tiers.

In resource migration typically the tasks are migrated from one resource to another resource in cloud. Hence it involves more of task migration [18] to gain better benefit of execution efficiency and performance leveraging powerful computing and efficient usage of storage capacity of clouds. During task migration there are conflicting objectives to be considered when making migration decisions, such as less energy consumption, and quick response in order to find an optimal migration path. A genetic algorithm based approach is used in effectively addressing the task migration problems. The genetic algorithm is used in formulating a decisionmaking for cloud service providers to migrate the tasks by defining policies for obtaining optimized task allocation schemes.

2.5 Resource Optimization

Achieving resource optimization among mobile users is a difficult task as each user doesn't have information about others who are using resources only cloud providers need to check the optimize usage of cloud resources. The resource optimization keeps track over and underutilization of resources by mobile users and prevent them wastage of resources so that they can be allotted to others who are in need of resources. It also keeps eye on price optimization so that the mobile users pays for only that resources for which he has used.

A novel framework was developed to model mobile applications as a local time workflows of tasks, where the mobility patterns are translated to a mobile service usage patterns. An efficient heuristic algorithm called MuSIC [19] evaluates optimized usage of cloud services using 2-tier mobile cloud approach by improving QoS.A distributed optimized algorithm [20] is developed to optimize the usage of cloud services by mobile devices and mobile cloud service providers runs separate VM's to execute the jobs for mobile device users by minimizing the electrical cost and maximizing their revenue cost.

The dynamic offloading problem [21]in context of MCC is provided by considering multi-dimensional way of accessing the cloud using availability of WLAN hotspots, energy consumption, communication costs and expected delays. A Markov decision process is used to derive near-optimal offloading policy. A deterministic delay constrained task partitioning [22] is used to solve offloading decision problem with delay constraints which are generalized by setting the dependency of tasks in a tree and a probabilistic delay constrained task partitioning is designed to guarantee within QoS constraints.A MAUI [23] is a system which uses fine grain code offload mechanism for mobile users so that they can offload part of code which is critical to cloud and execute it with maximum energy savings.

3 Conclusion and Future Research Directions

This paper highlights the different issues of resource provisioning and provides solutions to those issues but still some issues have uncertain problems like dynamic scaling of load balancers, dynamically managing the network in synchrony with VM provisioning, Achieving replication without increasing in high performance degradation, generalizing price prediction mechanisms and cloud services reactive location strategies.

5 References

[1] White Paper, "Mobile Cloud Computing Solution Brief," AEPONA, November 2010.

[2] SDSM: A Secure Data Service Mechanism in Mobile Cloud Computing 978-1-4577-0248-8/11/$26.00 ©2011 IEEE

[3] Jacson H. Christensen, "Using RESTful web-services and cloud computing to create next generation mobile applications," in Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications (OOPSLA), pp. 627-634, October 2009

[4] J. Oberheide , K. Veeraraghavan, E. Cooke, J. Flinn, and F. Jahanian. "Virtualized in-cloud security services for mobile devices," in Proc 1st Workshop on Virtualization in Mobile Computing (MobiVirt), pp. 31-35, June 2008.

[5] M. Reza Rahimi,, Jian Ren ,, Chi Harold Liu,, Athanasios V. Vasilakos,,Nalini Venkatasubramanian,"Mobile Cloud Computing: A Survey, State of Art and Future Directions**",Article in** Mobile Networks and Applications ,April 2014, Volume 19, Issue 2, pp 133-143.

[6] O. A. Ben-Yehuda, M. Ben-Yehuda, A. Schuster, and D. Tsafrir, "The Resource-as-a-Service (RaaS) cloud," in HotCloud '11: 4th USENIX Workshop on Hot Topics in Cloud Computing, 2012.

[7] M. Reza. Rahimi, N. Venkatasubramania "MAPCloud: Mobile Applications on an Elastice 2-Tier Cloud Architecture", submitted to IEEE GLOBECOM 2012.

[8] Z. Cai, X. Li and J.N.D. Gupta, "Heuristics for Provisioning Services to Workflows in XaaS Clouds", *IEEE Transactions in Services Computing*, no. 99, pp. 1

[9] H. Shen and G. Liu, "An Efficient and Trustworthy Resource Sharing Platform for Collaborative Cloud Computing", *IEEE Trans. Parallel Distrib. Sys.*,

[10] DusitNiyato, Ping Wang, EkramHossainWaliSaad, and Zhu Han, "Game Theoretic Modeling of Service Providers in Mobile Cloud Computing Environments",IEEE wireless Communications and Networking Conference:services,applications.and Business ,978-1-4673-0437-5,Dec 2012.

[11] C. Xian, Y. Lu, and Z. Li, "Adaptive computation offloading for energy conservation on battery-powered systems," in *IEEE International Conference on Parallel and Distributed Systems (ICPAD)*, vol. 2, pp. 1–8, 2007.

[12] Hongbin Liang, Tianyi Xing, Lin X. Cai, Dijiang Huang, Daiyuan Peng, and Yan Liu, "Adaptive ComputingResource Allocation for Mobile Cloud Computing", International Journal of Distributed Sensor Networks Volume 2013, Article ID b 81426, 14 pages http://dx.doi.org/10.1155/2013/181426

[13] MudassarAhmad,"Prognostic Load Balancing Strategy for Latency Reduction in Mobile Cloud Computing", Middle-East Journal of Scientific Research, vol.6, pp.805-813, 1990.

[14] WuqiGao et al. [2012], "Cloud simulation resource scheduling algorithm based on multi-dimensional quality of service", Information Technology journal-Asian network for scientific information. Pp. 99-101, ISSN: 1812-5638.

[15] Naela Rizvi, Prashant Pranav, Bibhav Raj, Sanchita Paul "Auction Model Using RR Approach for SLA Based Resource Provisioning in Multi-Cloud Environment", International Journal of Engineering and TechnologyVolume 8 Issue 2 pp 774-784 e-ISSN : 0975-4024 p-ISSN : 2319-8613

[16] G. Yalcin, O.S. Unsal, A. Cristal and M. Valero, "FIMSIM: A fault injection infrastructure for microarchitectural simulators", *ICCD*, 2011

[17] Leonardo P. Tizzei, M. A. S. Netto, Shu Tao "Optimizing Multi-tenant Cloud Resource Pools via Allocation of Reusable Time Slots"12th International Conference on Economics of Grids, Clouds, Systems and Services*, 2015*
[18] V. Cardellini, "A game-theoretic approach to computation offloading in mobile cloud computing", *Math. Program.*, pp. 1-29, 2015
[19] Zhang, W., Tan, S., Lu, Q., Liu, X., & Gong, W. (2015). A Genetic-Algorithm-Based Approach for Task Migration in Pervasive Clouds. International Journal of Distributed Sensor Networks..
[20] M. R. Rahimi, N. Venkatasubramanian and A. V. Vasilakos, "MuSIC: Mobility-aware optimal service allocation in mobile cloud computing", *Proc. IEEE Int. Conf. Cloud Comput.*, pp. 75-82, 2013
[21] Chunlin LI , Layuan LI ," An Optimization Approach for Utilizing Cloud Services for Mobile Devices in Cloud Environment, INFORMATICA, 2015, Vol. 26, No. 1, 89–110 89@2015 Vilnius University DOI: http://dx.doi.org/10.15388/Informatica.2015.40
[22] S. Kosta, A. Aucinas, P. Hui, R. Mortier and X. Zhang, "ThinkAir: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading", *Proc. IEEE INFOCOM '12*, pp. 945-953
[23] Angin, P., Bhargava, B.: An agent-based optimization framework for mobile-cloud computing. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA) 4, 1–17 (2013)