RESOURCE ALLOCATION IN MOBILE CLOUD COMPUTING: A SURVEY

Mr. C.Arun[1], Dr. K.Prabu[2]

[1]Research Scholar, PG & Research Department of Computer Science, Sudharsan College of Arts & Science, Pudukkottai, Tamilnadu, India.
[2]Assistant Professor, PG & Research Department of Computer Science, Sudharsan College of Arts & Science, Pudukkottai, Tamilnadu, India.

ABSTRACT- *Mobile Cloud Computing (MCC) has been introduced as a possible solution to the inherited limitations of mobile computing. The combination of wireless communication, cloud computing and portable computing devices is called mobile cloud computing which permits users an online access. By using MCC, the processing and the storage of intensive mobile device jobs will take place in the cloud system and the results will be returned to the mobile device. But the mobile cloud computing have some issues resource poverty, like power consumption, and security. Using the mobile devices for accessing the cloud it needs an efficient dynamic resource allocation for allocating the resources to the users.*

Key words: Mobile device, Mobile computing, Cloud computing, Power consumption, Resource poverty, Security.

1. INTRODUCTION

Mobile communication is the process of performing computations on a mobile device and transmission of data to and from one or more devices [1]. It is the method of getting connected and making use of centrally located information and application software with the deployment of small, portable wireless communication and computing devices [1–3]. Mobile communication facilitates the execution of a number of applications on a single device. In this modernistic world, everything is exceedingly dependent on technology. With the increase in the number of mobile users day by day, the need to provide better quality of service at very little power and cost also increases. At the same time cloud computing is a development in the field of computer science and technology. Computer users access Internet services via lightweight movable devices because powerful desktop machines are going through a phase of drought. Cloud is a distributed computing paradigm. It is a collection of interconnected and virtualized computers, which are provisioned and presented dynamically as unified computing resources offered on a pay-per-use basis [4]. Cloud computing is defined as applications that are delivered as Internet services: the hardware and system software in the data centers are used to provide these services [5]. Cloud computing is an advanced technology that focuses on the way of developing applications, designing computing systems and leveraging existing services for building software [6]. It is based on dynamic provisioning. In cloud computing, resources are offered in an on-demand and pay-per-use basis from the cloud computing vendors. So, integration of mobile computing with cloud computing (CC) has given birth to a newer and better technological approach called mobile cloud computing (MCC), In a simple sense, MCC is nothing but cloud computing in which mobile devices are involved as the thin clients. Here, data will be offloaded into cloud from mobile devices for computation or storage.

2. RESOURCE MANAGEMENT

In computing, resource allocation is necessary for any application to be run on the system. When the user opens any program this will be counted as a process, and therefore requires the computer to allocate certain resources for it to be able to run. Provisioning satisfies the request by mapping virtualized resources to physical ones. The hardware and software resources are allocated to the cloud applications on-demand basis. For scalable computing, Virtual Machines are rented. Lack of consciousness in resource management is another reason for excessive energy consumption. To overcome this, the resource management system should be improved for both cloud servers and mobile devices [7]. The resource management system must be such that resources are allocated in an on-demand fashion as the process progresses its execution in Figure 1. There should be no provisioning of allocation of resources prior to execution since it is difficult to correctly predict the resource needs before execution and unnecessary resource utilization may happen with allocation before the execution begins.

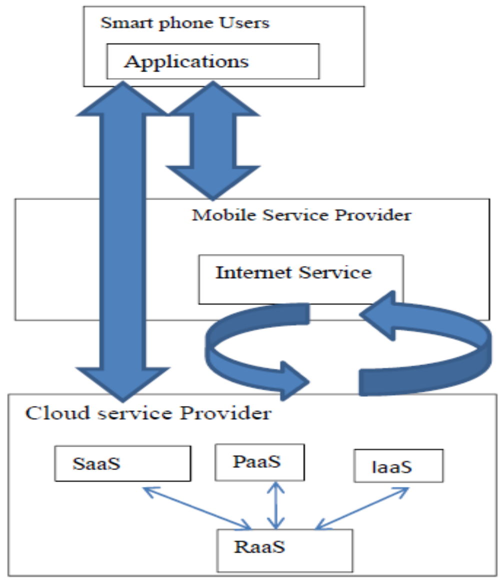

Fig -1: Resource allocation in MCC

3. SIGNIFICANCE OF RESOURCE ALLOCATION IN MOBILE CLOUD COMPUTING

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 such as availability, throughput, security, response time, reliability and performance.

There is a huge significance and issues of resource allocation [8] in MCC. Only an appropriate resource allocation can augment mobile devices and fulfill users' demand in Figure 2. Thus, the cloud maintains its QoS. Without a resource-allocation strategy, many diverse situations may occur. An optimal resource-allocation strategy can avoid the following situations:

1. Scarcity of resources: This situation arises when there are limited resources.
2. Resource contention: This situation arises when two applications try to access the same resource at the same time.
3. Resource fragmentation: This situation arises when resources are isolated. Despite being abundant, resources cannot be allocated to the needed application.
4. Over-provisioning: This situation arises when the application receives more resources than it demanded.
5. Under-provisioning: This situation occurs when the application is assigned with fewer resources than its demands.

Hence, resource allocation plays a significant role in MCC.

Fig -2: Resource provisioning issues in MCC

Y.Liu et. al., [9] proposed a multi-resource allocation strategy enhances to the quality of mobile cloud service, in terms of the system throughput and the service latency also frame the resource allocation model as a semi-Markov decision process under the average cost criterion. To solve the optimization problem, linear programming technology was used. An optimal resource allocation policy is calculated by the system requirements of the request blocking probability and service time latency.

To improves the QoS in mobile devices by choosing suitable cloud providers for a given user application M.H.Zarei et. al., [10] formulated the task allocation problem as a minimum nonlinear optimization problem. A randomized algorithm is proposed in which the result is close to the optimal solution. The proposed algorithm is faster and more scalable compared to optimal solution.

S.Vakilinia et. al., [11] proposed a model for wireless interfaces, mobile application profiles and cloud resources. First, an algorithm to allocate wireless interfaces and cloud resources has been introduced. The proposed model is based on the wireless network cloud (WNC) concept. Then, considering power consumption, application quality of service (QoS) profiles, and corresponding cost functions, a multi-objective optimization approach using an event-based finite state model and dynamic constraint programming method has been used to determine the appropriate transmission power, process power, cloud offloading and optimum QoS profiles. The proposed algorithm save the mobile battery life and guarantees both QoS and cost simultaneously.

C.Li et. al., [12] proposed a hybrid mobile cloud computing system, in which mobile applications can use different resources , services in local cloud and remote public cloud such as storage, computation and bandwidth. The cross-layer load-balancing based mobile cloud resource allocation optimization is proposed. The proposed method augments local cloud service pools with public cloud to increase the probability of meeting the service level agreements. The problem is divided by public cloud service allocation and local cloud service allocation, which is achieved by public cloud supplier, local cloud agent and the mobile user. The system status information is used in the hybrid mobile cloud computing system such as energy, the preferences of mobile applications, server load in cloud datacenter to improve resource utilization and quality of experience of mobile user. Therefore, the system status of hybrid mobile cloud is monitored continuously. The system design of load-balancing based cross-layer mobile cloud resource allocation is also proposed.

S.Kuribayashi et. al., [13] proposed to enhance the existing joint multiple resource allocation method, so as to provide the following two functions: (1)a function to take account of the total processing time of network delay and service processing time in allocating resources. (2) a function to prevent the degradation in service quality of other request types when requests that require a short network delay occur more than expected.

Y.Zhang et. al., [14] presented the resource allocation process of a mobile cloud computing system as an auction mechanism with premium and discount factors. The premium and discount factors indicate complementary and substitutable relations among cloud resources provided by the service provider. Then, they analyze the individual rationality and incentive compatibility (truthfulness) properties of the users in the proposed auction mechanism.

S.Falavarjani et. al., [15] proposed a two-stage method to solve the problem: first, NSGA-II was applied to obtain the Pareto solution set; second, entropy weight and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) technique are employed to stipulate the best compromise solution. A context-aware offloading middleware was developed to collect contextual information and handle offloading process. To stimulate selfish users, a virtual credit based incentive mechanism was exploited in offloading decision. The ability of the proposed resource allocation approach was to manage the trade-off between time and energy.

P.P.Hung et. al., [16] presented a comprehensive thin-thick client collaboration that involves conventional desktop or laptop computers, known as thick clients, by allowing the thin client to borrow resources from thick clients, particularly for optimizing data distribution and utilizing MCC resources to meet Service- Level Agreements, Quality-of-Service requirements and cloud service customers' budget. The proposed architecture was improved resource allocation efficiency and achieve better performance.

H.Liang et. al., [17] proposed a novel MCC adaptive resource allocation model to achieve the optimal resource allocation in terms of the maximal overall system reward by considering both mobile devices and cloud. To achieve this, the adaptive resource allocation as a semi-Markov decision process (SMDP) to capture the dynamic arrivals and departures of resource requests. The proposed model can achieve higher system reward and lower service blocking probability compared to traditional approaches based on greedy resource allocation algorithm.

J.Wei et. al., [18] proposed a novel resource allocation approach AIMING, which aims to minimize the latency constrained by monetary overhead in the context of federated-cloud. The network resources are deployed and selected according to k-means clustering. The total latency among datacenters is optimized based on binary quadratic programming. The evaluation is conducted with real data traces. The proposed AIMING was reduced total datacenter latency effectively compared with other approaches.

W. Chen et. al., [19] introduced a heuristic algorithm to solve NP-hard problem. When no infrastructure is available, compared with the case when an application is executed solely on a single MD, the proposed algorithm was reduced an application's response time significantly.

H.Raei et. al., [20] analyzed completely modeled three common resource allocation scheme (RAS), namely share-based scheme (SBS), reserve-based scheme (RBS), and hybrid-based scheme (HBS). The proposed models enable the cloudlet owner to properly decide which scheme is suitable for its conditions. The principal criteria for this decision are two important performance measures: request rejection probability and mean response delay. To model each scheme, an analytical performance model which consists of stochastic sub-models was proposed. The Markov Reward Model (MRM) is applied for obtaining the outputs of the sub-models.

L.Tang et. al., [21] proposed and analyzed a double-sided bidding mechanism where each demanding user submits a bid to choose a demand resource-price function and each supplying user submits a bid to choose a supply resource price function. They consider the cases with price-taking users and price-anticipating users who would anticipate the influence of their own bids on the price, respectively. For the case with price-taking users, they first proposed mechanism admits a unique competitive equilibrium which maximizes the social welfare of the mobile cloud, and then develop an optimal distributed algorithm to achieve the desired equilibrium point. For the case with price-anticipating users, they first formulate the interaction among multiple users as a strategic game and prove the existence and uniqueness of Nash equilibrium, and then develop a distributed algorithm to compute the Nash equilibrium. The proposed algorithms and illustrate that the social welfare achieved at Nash equilibrium is very close to the optimal social welfare.

A.Jin et. al., [22] proposed a feasible and truthful incentive mechanism (TIM), to coordinate the resource auction between mobile devices as service users (buyers) and cloudlets as service providers (sellers). Further, TIM is extended to a more efficient design of auction (EDA). TIM guarantees strong truthfulness for both buyers and sellers, while EDA achieves fairly high system efficiency but only satisfies strong truthfulness for sellers. They also show the difficulties for the buyers to manipulate the resource auction in EDA and the high expected utility with truthful bidding.

R. Agarwal et. al., [23] proposed a scheme that constructs a high-performance decentralized system by a group of volunteer mobile devices which come together to form a resourceful unit (cloudlet). The idea is to design a model to operate as a public resource between mobile devices in close geographical proximity. This cloudlet can provide larger storage capability and can be used as a computational resource by other devices in the network. The system needs to watch the movement of the participating nodes and restructure the topology if some nodes that are providing support to the cloudlet fail or move out of the network. To achieve this by leveraging the concept of virtual dominating set to create an overlay in the broads of the network and distribute the responsibilities in hosting a cloudlet server. They also proposed architecture called DRAP, for such a system and develop algorithms that are requited for its operation.

To overcome the main bottlenecks of wireless bandwidth between mobile terminal and cloudlet, and the computation capability of cloudlet, the joint optimization strategy is S.Meng et. al., [24] proposed to enhance the quality of mobile cloud service to formulate the wireless bandwidth and computing resource allocation model as a triple-stage Stackelberg game, and solve it by using backward method. In addition, the interplays of triple-stage game are discussed and the sub game optimal equilibrium for each stage is analyzed. An iterative algorithm is proposed to obtain Stackelberg equilibrium. Numerical results demonstrate the effectiveness of the proposed algorithm.

P.Akk et. al., [25] proposed an entropy based FIFO method to allocate resources to the mobile devices.

Rakesh Kumar et. al., [26] proposed a model to considered the feedback sent by the mobile device and stored it in a directory maintained by the root server at cloud. The root server refers this directory while allocating the task to cloudlets. For the second problem they used the Gabriel architecture and crowd sensing framework collectively. The combination of these two quite efficiently processes the sensed information at the local level and passes the processed information to root server for decision making. They also proposed metrics of yield factor for various parameters which will be calculated by the root server and based on those yield factor values root server can decide whether to scale up the cloud–cloudlet system or not.

S. Durga et. al., [27] proposed model to provide better quality of service to the mobile cloud customers. A Cuckoo based allocation strategy is by and the allocation is considered as an optimization problem with the aim of reducing the make span and the computational cost meeting the deadline constraints, with high resource utilization.

4. CONCLUSION

Resource allocation is one of the main operational issues in an MCC environment. After the task is offloaded from the mobile device into the cloud, the cloud provider allocates the task with the desired resources. In this paper, we discussed about the significance and we also discussed about previous of resource allocation models and different approaches. We demand to adapt the difficulties of the present approaches and overcome all the challenges of resource-allocation strategies to maintain proper QoS of the MCC system.

REFFERENCES

[1]. R. Kamal, Mobile Computing, Oxford University Press, Inc., Oxford, U.K., 2008.

[2]. T. S. Rappaport, Wireless Communications: Principles and Practice, vol. 2, Prentice Hall PTR, Upper Saddle River, NJ, 1996.

[3]. L. S. Ashiho, Mobile technology: Evolution from 1G to 4G, Electronics for You, 94–98, 2003.

[4]. R. Buyya, J. Broberg, and A. M. Goscinski, Cloud Computing: Principles and Paradigms, John Wiley & Sons, Hoboken, NJ, 2010.

[5]. B. Sosinsky, Cloud Computing Bible, John Wiley & Sons, Indianapolis, IN, 2010.

[6]. R. Buyya, C. Vecchiola, and S. T. Selvi, Mastering Cloud Computing, Tata McGraw-Hill Education, New Delhi, India, 2013.

[7]. M. Rahman, J. Gao, and W. Tsai, Energy saving in mobile cloud computing, in Proceedings of IEEE International Conference on Cloud Engineering, Redwood City, CA, pp. 285–291, 2013.

[8]. R. Patel and S. Patel, Survey on resource allocation strategies in cloud computing, International Journal of Engineering Research & Technology, 2(2), 1–5, 2013.

[9]. Y.Liu and M.J. Lee," Security-Aware Resource Allocation for Mobile Cloud Computing Systems", DOI: 10.1109/ICCCN.2015.7288465

[10]. M.H.Zarei and M.A. Shirsavar," A QoS-aware Task Allocation Model for Mobile Cloud Computing", DOI: 10.1109/ICWR.2016.7498444

[11]. S.Vakilinia, D.Qiu and M.Ali," Optimal multi-dimensional dynamic resource allocation in mobile cloud computing", EURASIP Journal on Wireless Communications and Networking 2014.

[12]. C. Li and L.Li," Load-Balancing Based Cross-Layer Elastic Resource Allocation in Mobile Cloud", DOI 10.1007/s11277-017-4615-3.

[13]. S.Kuribayashi," Resource Allocation Method For Cloud Computing Environments With Different Service Quality To Users At Multiple Access Points", International Journal of Computer Networks & Communications (IJCNC) Vol.7, No.6, November 2015. http:// DOI: 10.5121/ijcnc.2015.7603 33.

[14]. Y.Zhang, D.Niyato, and P.Wang," An Auction Mechanism for Resource Allocation in Mobile Cloud Computing Systems", https://doi.org/10.1007/978-3-642-39701-1_7

[15]. S.G.Falavarjani, M.Nematbakhsh and B.Ghahfarokhi," Context-aware multi-objective resource allocation in mobile cloud", https://doi.org/10.1016/j.compeleceng.2015.02.006

[16]. P. P. Hung, T.A.Bui, K.Soonil and E.N.Huh," A New Technique for Optimizing Resource Allocation and Data Distribution in Mobile Cloud Computing", http://dx.doi.org/10.5755/j01.eee.22.1.14113.

[17]. H.Liang, T. Xing, L.X. Cai, D.Huang, D.Peng, and Y.Liu," Adaptive Computing Resource Allocation for Mobile Cloud Computing", http://dx.doi.org/10.1155/2013/181426.

[18]. J.Wei , A.Zhou, J.Yuan, and F. Yang," AIMING: Resource Allocation with Latency Awareness for Federated-Cloud-Applications", http://doi.org/10.1155/2018/4593208.

[19]. W.w.Chen, C.Lea and K.Li," Dynamic Resource Allocation in Ad-Hoc Mobile Cloud Computing", DOI: 10.1109/WCNC.2017.7925613

[20]. H.Raei and N.Yazdani," Analytical performance models for resource allocation schemes of cloudlet in mobile cloud computing", DOI 10.1007/s11227-016-1830-8

[21]. L.Tang, S.He and Q.Li," Double-sided Bidding Mechanism for Resource Sharing in Mobile Cloud", DOI 10.1109/TVT.2016.2565505.

[22]. A.Jin, W.Song, P.Wang and P.Ju," Auction Mechanisms Toward Efficient Resource Sharing for Cloudlets in Mobile Cloud Computing", DOI 10.1109/TSC.2015.2430315

[23]. Zadhika Agarwal, Amiya Nayak,"DRAP: A Decentralized Public Resourced Cloudlet for Ad-hoc Networks", DOI: 10.1109/CloudNet.2015.7335327

[24]. S.Meng, Y.Wang, Z.Miao and K.Sun", Joint optimization of wireless bandwidth and computing resource in cloudlet-based mobile cloud computing Environment", DOI 10.1007/s12083-017-0544-x.'

[25]. P.Akki and Y.M.ROOPA", Resource Allocation Using Entropy Based Fifo Method In Mobile Cloud Computing ", International Journal of Engineering Sciences Research-IJESR, Vol 04, Special Issue 01, 2013.

[26]. R. Kumar and S. Yadav," Scalable Key Parameter Yield of Resources Model for Performance Enhancement in Mobile Cloud Computing", DOI 10.1007/s11277-017-4035-4

[27]. S. Durga, S. Mohan, J. Dinesh," Cuckoo Based Resource Allocation for Mobile Cloud Environments", https://doi.org/10.1007/978-981-10-0251-9_50.